PATENT COOPERATION TREATY **09/701868**

From the INTERNATIONAL SEARCHING AUTHORITY

To: JANELLE S. GRAETER
U.S. DEPARTMENT OF AGRICULTURE ARS-OTT
5601 SUNNYSIDE AVENUE
ROOM-4-1186
BELTSVILLE, MARYLAND 20705-5131

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT OR THE DECLARATION

(PCT Rule 44.1)

	Date of Mailing *(day/month/year)* 03 NOV 1999
Applicant's or agent's file reference PPD50352 PCT	**FOR FURTHER ACTION** **See paragraphs 1 and 4 below**
International application No. PCT/US99/12697	International filing date *(day/month/year)* 08 JUNE 1999
Applicant U.S. DEPARTMENT OF AGRICULTURE	

1. [X] The applicant is hereby notified that the international search report has been established and is transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally 2 months from the date of transmittal of the international search report; however, for more details, see the notes on the accompanying sheet.

   Where? Directly to the International Bureau of WIPO
   34, chemin des Colombettes
   1211 Geneva 20, Switzerland
   Facsimile No.: (41-22) 740.14.35

   For more detailed instructions, see the notes on the accompanying sheet.

2. [ ] The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. [ ] **With regard to the protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

   [ ] the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

   [ ] no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Further action(s):** The applicant is reminded of the following:

   Shortly after **18 months** from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in rules 90 *bis* 1 and 90 *bis* 3, respectively, before the completion of the technical preparations for international publication.

   Within **19 months** from the priority date, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later).

   Within **20 months** from the priority date, the applicant must perform the prescribed acts for entry into the national phase before all designated Offices which have not been elected in the demand or in a later election within 19 months from the priority date or could not be elected because they are not bound by Chapter II.

Name and mailing address of the ISA/US	Authorized officer	
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	MELISSA KIMBALL	JOYCE BRIDGERS PARALEGAL SPECIALIST CHEMICAL MATRIX

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference PPD50352 PCT	FOR FURTHER ACTION	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
International application No. PCT/US99/12697	International filing date *(day/month/year)* 08 JUNE 1999	(Earliest) Priority Date *(day/month/year)* 09 JUNE 1998
Applicant U.S. DEPARTMENT OF AGRICULTURE		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 4 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. [X] **Certain claims were found unsearchable** (See Box I).

2. [ ] **Unity of invention is lacking** (See Box II).

3. [ ] The international application contains disclosure of a **nucleotide and/or amino acid sequence listing** and the international search was carried out on the basis of the sequence listing
   - [ ] filed with the international application.
   - [ ] furnished by the applicant separately from the international application,
     - [ ] but not accompanied by a statement to the effect that it did not include matter going beyond the disclosure in the international application as filed.
   - [ ] transcribed by this Authority.

4. With regard to the **title**,
   - [X] the text is approved as submitted by the applicant.
   - [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,
   - [X] the text is approved as submitted by the applicant.
   - [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is:

   Figure No. ______
   - [ ] as suggested by the applicant.
   - [ ] because the applicant failed to suggest a figure.
   - [ ] because this figure better characterizes the invention.
   - [X] None of the figures.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12697

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-26 and 28-32
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

because the claims all recite SEQ ID No.s or depend therefrom while no CRF has been filed for this case. Therefore it is not possible to search the claimed nucleic acid and amino acids nor is it possible to search transgenic seeds or plants comprising the sequences.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12697

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12N 5/04, 9/38, 15/09, 15/56; A01H 5/00, 5/10
US CL : 435/207, 419, 468; 800/278, 295, 298
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/207, 419, 468; 800/278, 295, 298

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST, CAPLUS, AGRICOLA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SMITH et al. A Gene Coding for Tomato Fruit β-Galactosidase II Is Expressed during Fruit Ripening. Plant Physiology. 1998, Vol. 117, pages 417-423, especially 422-423.	27
Y	ALI et al. Isolation, Characterization and Significance of Papaya β-Galactanases to Cell Wall Modification and Fruit Softening during Ripening. Physiologia Plantarum. 1998, Vol. 104, pages 105-115, especially page 111, col. 2, and page 113, col. 2.	27

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 OCTOBER 1999	03 NOV 1999

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks	JOYCE BRIDGERS PARALEGAL SPECIALIST

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12697

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CARRINGTON et al. β-Galactosidase II Activity in Relation to Changes in Cell Wall Galactosyl Composition during Tomato Ripening. Journal of the American Society of Horticultural Science. 1996, Vol. 121, No. 1, pages 132-136, especially page 135, col. 2.	27
Y	PRESSEY, R. β-Galactosidases in Ripening Tomatoes. Plant Physiology. 1983, Vol. 71, pages 132-135, see entire article.	27
Y,P	US 5,859,344 A (BIRD et al.) 12 January 1999, see entire document.	27

09/701868

PATENT COOPERATION TREATY

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To: JANELLE S. GRAETER
U.S. DEPARTMENT OF AGRICULTURE ARS-OTT
5601 SUNNYSIDE AVENUE
ROOM-4-1186
BELTSVILLE, MARYLAND 20705-5131

PCT

NOTIFICATION OF TRANSMITTAL OF INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Rule 71.1)

Date of Mailing (*day/month/year*)	27 NOV 2000

Applicant's or agent's file reference	IMPORTANT NOTIFICATION
PPD50352 PCT	

International application No.	International filing date (*day/month/year*)	Priority Date (*day/month/year*)
PCT/US99/12697	08 JUNE 1999	09 JUNE 1998

Applicant
U.S. DEPARTMENT OF AGRICULTURE

1. The applicant is hereby notified that this International Preliminary Examining Authority transmits herewith the international preliminary examination report and its annexes, if any, established on the international application.

2. A copy of the report and its annexes, if any, is being transmitted to the International Bureau for communication to all the elected Offices.

3. Where required by any of the elected Offices, the International Bureau will prepare an English translation of the report (but not of any annexes) and will transmit such translation to those Offices.

4. **REMINDER**

   The applicant must enter the national phase before each elected Office by performing certain acts (filing translations and paying national fees) within 30 months from the priority date (or later in some Offices)(Article 39(1))(see also the reminder sent by the International Bureau with Form PCT/IB/301).

   Where a translation of the international application must be furnished to an elected Office, that translation must contain a translation of any annexes to the international preliminary examination report. It is the applicant's responsibility to prepare and furnish such translation directly to each elected Office concerned.

   For further details on the applicable time limits and requirements of the elected Offices, see Volume II of the PCT Applicant's Guide.

Name and mailing address of the IPEA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	MELISSA KIMBALL Joyce Bridgers
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/IPEA/416 (July 1992)★

09/701868

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference PPD50352 PCT	**FOR FURTHER ACTION** See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/US99/12697	International filing date (*day/month/year*) 08 JUNE 1999	Priority date (*day/month/year*) 09 JUNE 1998
International Patent Classification (IPC) or national classification and IPC Please See Supplemental Sheet.		
Applicant U.S. DEPARTMENT OF AGRICULTURE		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 5 sheets.

   ☐ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority. (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 0 sheets.

3. This report contains indications relating to the following items:

   I ☒ Basis of the report
   II ☐ Priority
   III ☒ Non-establishment of report with regard to novelty, inventive step or industrial applicability
   IV ☐ Lack of unity of invention
   V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
   VI ☐ Certain documents cited
   VII ☐ Certain defects in the international application
   VIII ☐ Certain observations on the international application

Date of submission of the demand 07 JANUARY 2000	Date of completion of this report 26 OCTOBER 2000
Name and mailing address of the IPEA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Joyce Bridgers MELISSA KIMBALL Telephone No. (703) 308-0196

Form PCT/IPEA/409 (cover sheet) (July 1998)*

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US99/12697

I. Basis of the report

1. With regard to the elements of the international application:*

[X] the international application as originally filed

[X] the description:
pages 1-43 , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of

[X] the claims:
pages 44-50 , as originally filed
pages NONE , as amended (together with any statement) under Article 19
pages NONE , filed with the demand
pages NONE , filed with the letter of

[X] the drawings:
pages 1-31 , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of

[X] the sequence listing part of the description:
pages NONE , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of

2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.
These elements were available or furnished to this Authority in the following language ________ which is:

[ ] the language of a translation furnished for the purposes of international search (under Rule 23.1(b)).

[ ] the language of publication of the international application (under Rule 48.3(b)).

[ ] the language of the translation furnished for the purposes of international preliminary examination (under Rules 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

[ ] contained in the international application in printed form.

[ ] filed together with the international application in computer readable form.

[ ] furnished subsequently to this Authority in written form.

[ ] furnished subsequently to this Authority in computer readable form.

[ ] The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

[ ] The statement that the information recorded in computer readable form is identical to the writen sequence listing has been furnished.

4. [X] The amendments have resulted in the cancellation of:

[X] the description, pages NONE

[X] the claims, Nos. NONE

[X] the drawings, sheets/~~fig~~ NONE

5. [X] This report has been drawn as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).**

* *Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17).*

***Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.*

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US99/12697

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been and will not be examined in respect of:

☐ the entire international application.

☒ claims Nos. 1-26 AND 28-32

because:

☐ the said international application, or the said claim Nos. _ relate to the following subject matter which does not require international preliminary examination *(specify)*.

☐ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. _ are so unclear that no meaningful opinion could be formed *(specify)*.

☐ the claims, or said claims Nos. _ are so inadequately supported by the description that no meaningful opinion could be formed.

☒ no international search report has been established for said claims Nos. 1-26 and 28-32 .

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

☐ the written form has not been furnished or does not comply with the standard.

☒ the computer readable form has not been furnished or does not comply with the standard.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US99/12697

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. statement

Novelty (N)	Claims	NONE	YES
	Claims	27	NO
Inventive Step (IS)	Claims	NONE	YES
	Claims	27	NO
Industrial Applicability (IA)	Claims	27	YES
	Claims	NONE	NO

2. citations and explanations (Rule 70.7)

Claim 27 lacks novelty under PCT Article 33(2) as being anticipated by Smith et al.

The claim is drawn to a method of modifying cell wall metabolism in a plant by expressing a DNA construct which modifies beta-galactosidase activity.

Smith et al. teach that beta-galactosidase is an enzyme active in modifying cell wall during fruit ripening in tomato (page 417, col. 1). They teach that they have cloned the cDNA that encodes beta-galactosidase II and that it is expressed during ripening (page 418, col. 1). Smith et al. teach that they have produced tomato plants comprising beta-galactosidase in the antisense orientation with *Agrobacterium*-mediated transformation (page 423, col. 1). This plant has modified beta-galactosidase activity due to the expression of the transgene.

Claim 27 lacks an inventive step under PCT Article 33(3) as being obvious over Smith et al. for the reasons above.

Claim 27 meets the criteria set out in PCT Article 33(4), because the method has industrial applicability in that it would be useful in producing fruits with modified ripening patterns.

------------------- NEW CITATIONS -------------------
NONE

Supplemental Box
(To be used when the space in any of the preceding boxes is not sufficient)

Continuation of: Boxes I - VIII

Sheet 10

CLASSIFICATION:

The International Patent Classification (IPC) and/or the National classification are as listed below:
IPC(7): C12N 5/04, 9/38, 15/09, 15/56; A01H 5/00, 5/10 and US Cl.: 435/207, 419, 468; 800/278, 295, 298

I. BASIS OF REPORT:

5. (Some) amendments are considered to go beyond the disclosure as filed:
NONE

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 5/04, 9/38, 15/09, 15/56, A01H 5/00, 5/10

A1

(11) International Publication Number: WO 99/64564

(43) International Publication Date: 16 December 1999 (16.12.99)

(21) International Application Number: PCT/US99/12697

(22) International Filing Date: 8 June 1999 (08.06.99)

(30) Priority Data:
60/088,805 9 June 1998 (09.06.98) US

(71) Applicant *(for all designated States except US)*: U.S. DEPARTMENT OF AGRICULTURE [US/US]; Room 4–1188, 5601 Sunnyside Avenue, Beltsville, MD 20705–5131 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GROSS, Kenneth, C. [US/US]; 4713 Knapp Court, Ellicott City, MD 21043 (US). SMITH, David, L. [US/US]; 9017 Lambskin Lane, Columbia, MD 21045 (US).

(74) Common Representative: U.S. DEPARTMENT OF AGRICULTURE; Graeter, Janelle, S. (ARS–OTT), Room 4–1186, 5601 Sunnyside Avenue, Beltsville, MD 20705–5131 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GENES CODING FOR TOMATO β–GALACTOSIDASE POLYPEPTIDES

(57) Abstract

A novel β–galactosidase gene family and DNA sequences derived from the cloning of cDNAs encoding products of these genes are provided, as exemplified by a β–galactosidase II protein which is encoded by a cDNA clone, pZBG2–1–4. A method for modifying cell wall metabolism which involves modifying the activity of at least one β–galactosidase, and thus modifying the quality of the fruit is also provided. Also provided by the present invention is a DNA construct including some or all of a β–galactosidase DNA sequence under control of a transcriptional initiation region operative in plants, so that the construct can generate RNA and, optionally, β–galactosidase polypeptide in plant cells. The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the recombinant vectors, as well as to methods of making such vectors and host cells and for using them for production of β–galactosidase polypeptides or peptides by recombinant techniques. The present invention also provides plant cells containing DNA constructs of the present invention; plants derived therefrom having modified β–galactosidase gene expression; and seeds produced from such plants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

GENES CODING FOR TOMATO β-GALACTOSIDASE POLYPEPTIDES

Field of the Invention

The present invention relates to a family of novel plant genes encoding polypeptides characterized by their ability to hydrolyze terminal non-reducing β-D-galactosyl residues from β-D-galactosides. More specifically, a polynucleotide sequence derived from a cDNA clone designated pZBG2-1-4 (referred to in U.S. Provisional Appln. No. 60/088,805 as pTomβgal 4), which encodes a specific plant polypeptide named β-galactosidase II, is provided. Also provided are cDNA clones encoding six other homologous polypeptides, methods of using these cDNA clones for producing β-D-galactoside polypeptides of the invention, and methods of modifying fruit quality by employment of a polynucleotide or polypeptide of the present invention.

Background of the Invention

The most conspicuous and important processes related to post-harvest quality of climacteric fruit are the changes in texture, color, taste, and aroma which occur during ripening. Because of the critical relationship that deleterious changes in texture have to quality and post-harvest shelf-life, emphasis has been placed on studying the mechanisms involved in the loss of firmness that occurs during tomato fruit ripening. Although fruit softening may involve changes in turgor pressure, anatomical characteristics and cell

wall integrity, it is generally assumed that cell wall disassembly leading to a loss of wall integrity is a critical feature. The most apparent changes, in terms of composition and size, occur in the pectic fraction of the cell wall (see references in Seymour and Gross, 1996).

Changes known to occur in the pectic fraction of the cell wall during fruit ripening include increased solubility, depolymerization, de-esterification and a significant net loss of neutral sugar containing side chains (Huber, 1983; Fischer and Bennett, 1991; Seymour and Gross, 1996). The best characterized pectin-modifying enzymes are polygalacturonase (endo-α1$\rightarrow$4-D-galacturonan hydrolase; E.C. 3.2.1.15; PG) and pectin methylesterase (E.C. 3.1.1.11; PME). Although PG and PME are relatively abundant and have substantial activity during tomato fruit ripening, softening still occurs, albeit with a slight delay, in fruit where PG (Smith *et al.* 1988, 1990) or PME (Tieman *et al.* 1992; Hall *et al.* 1993) gene expression and enzyme activity was significantly down-regulated in transgenic plants. Moreover, over-expression of PG in non-ripening mutant *rin* tomato fruit did not result in softening even though depolymerization and solubilization of pectin was evident (Giovannoni *et al.*, 1989).

Among the other known pectin modifications that occur during fruit development, one of the best characterized is the significant net loss of galactosyl residues which occurs in the cell walls of many ripening fruit (Gross and Sams, 1984; Seymour and Gross, 1996). Although some loss of galactosyl residues could result indirectly from the action of PG, β-galactosidase (exo-β(1$\rightarrow$4)-D-galactopyranoside; E.C. 3.2.1.23) is the only enzyme identified in

higher plants capable of directly cleaving β(1→4)galactan bonds, and probably plays a role in galactan sidechain loss (DeVeau *et al.*, 1993; Carey *et al.*, 1995; Carrington and Pressey, 1996). No endo-acting galactanase has yet been identified in higher plants. The view that β-galactosidase is active in releasing galactosyl residues from the cell wall during ripening is supported by the dramatic increase in free galactose, a product of β-galactosidase activity (Gross, 1984) and a concomitant increase in activity of a particular enzyme, designated β-galactosidase II, in tomatoes during ripening (Carey *et al.*, 1995). β-galactosidase activity is thought to be important in cell wall metabolism (Carey et al., 1995). β-Galactosidases are generally assayed using artificial substrates such as *p*-nitrophenyl-β-D-galactopyranoside (PNP), 4-methylumbelliferyl-β-D-galactopyranoside and 5-bromo-4-chloro-3-indoxyl-β-D-galactopyranoside (X-GAL). However, it is clear that β-galactosidase II is also active against natural substrates, *i.e.*, β (1→4)galactan (Carey *et al.*, 1995; Carrington and Pressey, 1996; Pressey, 1983). β-Galactosidase proteins have been purified and characterized in a number of other fruits including kiwifruits (Ross *et al.*, 1993), coffee (Golden *et al.*, 1993), persimmon (Kang *et al.*, 1994), and apple (Ross *et al.*, 1994).

Carey *et al.* (1995) were able to purify three previously identified β-galactosidases from ripening tomato fruit (Pressey, 1983), but only one (β-galactosidase II) was active against β(1→4)galactan. Even though they were able to identify putative β-galactosidase cDNA clones, none of the cDNA's deduced amino acid sequences matched the amino terminal sequence of the β-galactosidase II protein. Although β-galactosidase II, a protein present in

tomato (*Lycopersicon esculentum* Mill.) fruit during ripening and capable of degrading tomato fruit galactan has been purified, cloning of the corresponding gene has been elusive.

The modification of plant gene expression has been achieved by several methods. The molecular biologist can choose from a range of known methods to decrease or increase gene expression or to alter the spatial or temporal expression of a particular gene. For example, the expression of either specific antisense RNA or partial (truncated) sense RNA has been utilized to reduce the expression of various target genes in plants (as reviewed by Bird and Ray, 1991, Biotechnology and Genetic-Engineering Reviews 9:207-227). These techniques involve the incorporation into the genome of the plant of a synthetic gene designed to express either antisense or sense RNA. They have been successfully used to down-regulate the expression of a range of individual genes involved in the development and ripening of tomato fruit (Gray et al, 1992, Plant Molecular Biology, i9:69-87). Methods to increase the expression of a target gene have also been developed. For example, additional genes designed to express RNA containing the complete coding region of the target gene may be incorporated into the genome of the plant to "over-express" the gene product. Various other methods to modify gene expression are known; for example, the use of alternative regulatory sequences. The complete disclosure of each of the references cited above is fully incorporated herein by reference.

The need therefore exists to clone a gene for β-galactosidase II and related polypeptides, and using known methods of modification of plant gene expression, thereby to provide methods for modifying quality of fruits,

particularly by modifying the cell wall, thereby directly affecting the ripening of the fruit.

Summary of the Invention

The present invention is based on the discovery of novel DNA sequences derived from cDNA clones from a family of genes encoding β-galactosidases. The phylogenic tree based on the shared amino acid sequence identities for the DNA sequences of the present invention is shown in Figure 1A,B. Five cDNA and two RT-PCR clones, designated herein as TBG1, TBG2, TBG3, TBG4, TBG5, TBG6, and TBG7 and having the nucleic acid sequences designated SEQ ID NOs 1-7, respectively as shown in Figure 2, were identified which had a high degree of shared sequence identity to other known β-galactosidases. The corresponding amino acid sequences are designated herein as SEQ ID NOs 8-16, respectively and are shown in Figure 2 and 3. The nucleotide sequences for SEQ ID NOs 1-7 are recorded in Gen Bank with the following respective Accessions Numbers:

SEQ ID NO:1	TGB1	AF023847	deposit Sept 10, 1997
SEQ ID NO:2	TGB2	AF154420	deposited May 19, 1999
SEQ ID NO: 3	TGB3	AF154421	deposited May 20, 1999
SEQ ID NO:4	TGB4	AF020390	deposited Aug 21, 1997
SEQ ID NO:5	TGB5	AF154423	deposited May 20, 1999
SEQ ID NO:6	TGB6	AF154424	deposited May 20, 1999
SEQ ID NO: 7	TGB7	AF154422	deposited May 20, 1999

Throughout the following discussion, wherever TBG4 is indicated in the description of the invention, it is to be understood that TBG1-3 and 5-7 are also to be included in that description, unless otherwise indicated.

A method of providing a DNA sequence of the invention, either by cloning a cDNA (for instance, pZBG2-1-4) that codes for a protein of the present invention, such as β-galactosidase II, or by deriving the DNA sequence from genomic DNA, or by synthesis of a DNA sequence ab initio using the cDNA sequence as a guide is also provided.

A method for modifying cell wall metabolism which involves modifying the activity of at least one galactosidase, and thus modifying the quality of the fruit is also provided.

Also provided by the present invention is a DNA construct including some or all of an exemplary β-galactosidase DNA sequence under control of a transcriptional initiation region operative in plants, so that the construct can generate RNA in plant cells.

Also discovered is an enhancer/promoter associated with expression of the genes encoding β-galactosidase.

The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the recombinant vectors, as well as to methods of making such vectors and host cells and for using them for production of β-galactosidase polypeptides or peptides by recombinant techniques.

The present invention also provides plant cells containing DNA constructs of the present invention; plants derived therefrom having modified β-galactosidase gene expression; and seeds produced from such plants.

The β-galactosidase II protein of the present invention has demonstrated enzyme activity in cell wall disassembly leading to loss of tissue integrity and fruit softening. The β-galactosidase II protein also may be involved in cell wall turnover, which could be involved in cell extension and/or expansion and therefore plant growth and development.

By hydrolyzing galactose from the cell wall, the enzyme may allow ripening to commence and/or progress, since galactose may be involved in stimulating ethylene production alone or in conjunction with unconjugated N-glycans.

The β-galactosidase of the invention may be involved in conversion of chloroplasts (green – chlorophyll) to chromoplasts (red – lycopene) during fruit ripening by degrading chloroplast membrane galactolipids.

The family of genes represented by the nucleotide sequences shown in Figure 2 is expected to code for a group of similar enzymes with the same type of hydrolytic activity but with different tissue and/or substrate specificity's or cellular compartmentation profiles.

The β-galactosidase II protein of the present invention as well as other proteins encoded in the nucleotide sequences shown in Figure 2 may be used for preparation of pectin and other cell wall derived polymers with lowered galactosyl content for use in biofilms and solutions (for example in

clarification of fruit juices) requiring lower or higher cross-linking or viscomertric properties.

The present invention also provides β-galactosidase enzymes for use as components of enzyme mixtures for protoplast isolation.

Brief Description of the Figures

Figure 1A and 1B shows a phylogenic tree based on shared amino acid sequence identity among tomato β-galactosidase clones TGB1-7 and other known plant β-galactosidase polypeptides.

Figure 2 shows cDNA sequences [SEQ ID NOs: 1-7, respectively] for the seven β-galactosidase genes of the invention: TGB1, TGB2, TGB3, TGB4, TGB5, TGB6, TGB7.

Figure 3 shows multiple sequence alignment of the deduced amino acid sequences of tomato fruit for cDNA clones TGB1, TGB2, TGB3, TGB4, TGB5, TGB6 and TGB7 [SEQ ID NOs: 8-16, respectively] and various plant β-galactosidase cDNA clones.

Figure 4 shows autoradiograph of northern blot analysis of TBG expression in various plant tissues (flowers, leaves, roots and stems).

Figure 5 shows Autoradiograph of northern blot analysis of TBG expression in fruit tissues at different stages of development.

Figure 6 shows autogradiograph of northern blot analysis of TBG expression in fruit tissues (mature green or turning stage fruit peel, outer pericarp, inner paricarp and locular).

Figure 7 shows autoradiograph of northern blot analysis of TBG expression in normal and mutant fruit tissues.

Figure 8 shows autoradiograph of northern blot analysis of TBG expression in response to ethylene treatment of mature green fruit tissues.

Figure 9 shows Western blot analysis of TBG4 expression by yeast.

Figure 10 shows detection of β-galactosidase activity from pZBG2-1-4 expression in *E. coli.*

Figure 11 A - E (1-4) shows the comparative results of texture measurements for fruit from tomato plants containing antisense constructs to suppress TBG4 mRNA and fruit from the parental line.

Figures 12A - B show Northern blot analysis of TBG4 expression in transgenic fruit containing TBG4 antisense construct.

Figure 13 shows a Binary construct used to transform plants and express TBG4 (pZBG2-1-4) in the antisense orientation.

Detailed Description

The following detailed description is directed to a preferred embodiment of the present invention and is intended as illustrative of each of other DNA sequences of the present invention.

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding β-galactosidase polypeptides, particularly a β-galactosidase II polypeptide having the amino acid sequence shown in Figure 2. The DNA sequence of the exemplary β-galactosidase II cDNA clone of the invention, which was determined from a cDNA clone, pZBG2-1-4, encoding β-galactosidase II, is recorded in GenBank as Accession Number AF020390. Not all β-galactosidases possess *in vitro* activity against extracted cell wall material via the release of galactose from wall polymers containing β(1→4)-D-galactan. The polypeptide expressed from the exemplary β-galactosidase II clone, pZBG2-1-4, has been shown to exhibit β-galactosidase activity and exogalactinase activity.

The exemplary β-galactosidase II protein of the present invention, as shown in Figure 2, shares sequence homology with the amino acid sequence deduced from β-galactosidase cDNA clones of TBG2-7 and cDNA clones of the fruits of asparagus (accession number P45582), apple (accession number P48981), and carnation (accession number Q00662), as well as with β-galactosidase cDNA clones of a previously published sequence of a tomato β-galactosidase cDNA clone designated pTomβgal1 (accession number P48980) isolated from ripe 'Ailsa Craig' fruit (Carey *et al.*, 1995). The ORF of the clone TBG1 disclosed herein by the inventors (accession number AF023847)

is nearly identical to the cDNA previously described by Carey et al. As shown in Figure 2, the shared deduced sequence identity is high among all the published plant β-galactosidases of the seven clones (TBG1-7) and the other plant β-galactosidases.

BLAST searches of the database also indicated significant shared sequence identity between domains of the plant β-galactosidases and mammalian and fungal β-galactosidases, however little share sequence identity was detected with bacterial β-galactosidases.

As shown in Figure 1, the shared amino acid identity of TBG1 and TBG3 was high. TBG4 was also very similar to both TBG1 and 3. The amino acid sequences of TBG2 and 7 were unique because several regions of amino acid insertions appear throughout their sequence (Figure 3).

Nucleic Acid Molecules

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using a PCR-based dideoxynucleotide terminator protocol and an ABI automated DNA sequencer (such as the Model 373 from Applied Biosystems, Inc., Foster City, CA), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least

about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

By "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides, and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U), where each thymidine deoxyribonucleotide (T) in the specified deoxyribonucleotide sequence is replaced by the ribonucleotide uridine (U).

Using the information provided herein, such as the exemplary nucleotide sequence shown in Figure 2 [SEQ ID NO: 4], a nucleic acid molecule of the present invention encoding a β-galactosidase II polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning cDNAs using mRNA as starting material. Illustrative of the invention, the nucleic acid molecule described in Figure 2 [SEQ ID NO: 4] was discovered in a cDNA library derived from breaker, turning and pink fruit pericarp from 'Rutgers' tomato plants.

The complete sequence of the cDNA insert of pZBG2-1-4 is accessible in the GenBank (no. AF020390) and is provided in Figure 2 [SEQ ID NO: 4]. The cDNA insert is 2532 nucleotides (nt) long and contains a single, long open reading frame (ORF) predicted to start with the first in-frame ATG at nt 64 and end with TAA at nt 2238. This ORF codes for a 79 kD protein 724 amino acids long. The deduced amino acid sequence of pZBG2-1-4 shared significant amino acid identity to all published plant β-galactosidase sequences in the database (Figure 1A,B). When the entire ORF of each β-galactosidase gene was compared to pZBG2-1-4, the shared sequence identity was about 64% for tomato pTomβgal 1 (P48980), about 67.6% for apple (P48981), about 63% for asparagus (P45582) and about 55% for carnation (Q00662). As one of ordinary skill would appreciate, due to the possibilities of sequencing errors discussed above, the actual complete β-galactosidase II polypeptide encoded by the deposited cDNA, which comprises about 724 amino acids, may be somewhat longer or shorter. More generally, the actual open reading frame may be anywhere in the range of ±20 amino acids, more likely in the range of ±10 amino acids, of that predicted from either the first methionine codon from the N-terminus shown in Figure 2 [SEQ ID NO: 4]. In any event, as discussed further below, the invention further provides polypeptides having various residues deleted from the N-terminus of the complete polypeptide, including polypeptides lacking one or more amino acids from the N-terminus of the β-galactosidase II polypeptide described herein.

Leader and Mature Sequences

Analysis of the deduced amino acid sequence of pZBG2-1-4 suggested a high probability for secretion based on the presence of a hydrophobic leader sequence, a leader sequence cleavage site and three possible N-glycosylation sites. The programs PSORT V6.4 (Nakai and Kanehisa, 1992, incorporated herein by reference) and SignalP V1.1 (Nielsen et al., 1997, incorporated herein by reference), were used to predict that the ORF contains a hydrophobic leader sequence that would be cleaved between the alanine and serine residues at positions 23 and 24 respectively, and that the mature polypeptide has an extracellular location. The mature polypeptide contains three possible N-glycosylation sites at asparagine numbers 282, 459 and 713, however the asparagine at position 713 is unlikely to be glycosylated due to the proline at position 714. The predicted molecular mass of the unglycosylated mature polypeptide was 75 kD with a pI of 8.9.

Accordingly, the amino acid sequence of the complete β-galactosidase II protein of the invention includes a leader sequence and a mature protein, as shown in Figure 3 [SEQ ID NO: 4]. More in particular, the present invention provides nucleic acid molecules encoding a mature form of the β-galactosidase II protein. Thus, according to the signal hypothesis, secreted proteins have a signal or secretory leader sequence which is cleaved from the complete polypeptide to produce a secreted "mature" form of the protein. In some cases, cleavage of a secreted protein is not entirely uniform, which results in two or more mature species of the protein. Further, it has long been known that the cleavage specificity of a secreted protein is ultimately determined by the

primary structure of the complete protein, that is, it is inherent in the amino acid sequence of the polypeptide. Therefore, the present invention provides a nucleotide sequence encoding the mature β-galactosidase II polypeptide having the amino acid sequence encoded by the cDNA shown in Figure 2 [SEQ ID NO: 4] and provided in GenBank (Accession No. AF20390). By the "mature β-galactosidase II polypeptide having the amino acid sequence encoded by the cDNA clone shown in Figure 2 [SEQ ID NO: 4] is meant the mature form(s) of the β-galactosidase II protein produced by expression in a plant cell of the complete open reading frame encoded by the cDNA sequence of the clone shown in Figure 2 [SEQ ID NO: 4] and provided in GenBank (Accession No. AF20390).

The exemplary β-galactosidase II cDNA of the present invention (TBG4) has been expressed in *E. coli* strain XLI blue MR (lacZ) (Stratagene, La Jolla, CA), as described hereinbelow (see Example).

Analysis of the deduced amino acid sequence of cDNA clones representing the other β-galactosidase genes of the invention also revealed open reading frames and, in some cases, suggested a high probability for secretion of the encoded proteins. All the full-length cDNA clones were predicted to have a signal sequence (Fig. 2). Using the two prediction programs SignalP and PSORT, TBG4 was predicted to be secreted by both programs. TBG1, 2 and 3 were predicted to have cleavable signal sequences by SignalP, but uncleavable signal sequences by PSORT. TBG7 was suggested to be targeted to the chloroplast by PSORT. Particular observations for each of the seven clones are as follows, based on the presence of a hydrophobic

leader predicted by the programs PSORT V6. and SignalP V1.1: TBG1: initiation codon at 306 [SEQ ID NO: 1], ORF = 835 amino acids [SEQ ID NO: 8], signal sequence at 1-24; TBG2: initiation codon not determined [SEQ ID NO: 2], ORF = 888 amino acids [SEQ ID NO: 9], signal sequence at 1-25; TBG3: initiation codon at 32 [SEQ ID NO: 3], ORF = 838 amino acids [SEQ ID NO: 10], signal sequence at 1-22; TBG5: initiation codon not determined [SEQ ID NO:5], ORF = 251 amino acids [SEQ ID NO: 12], signal sequence not determined; TBG6: initiation codon not determined [SEQ ID NO:6], ORF = 248 amino acids [SEQ ID NO:13], signal sequence not determined; TBG7: initiation codon at 104 [SEQ ID NO: 7], ORF = 870 amino acids [SEQ ID NO:14], signal sequence at 1-35.

The deduced amino acid sequences of the seven clones was also subjected to analysis using the program DNAsis and the predictions for molecular mass, cellular targeting, pI and potential N-linked glycosylation sites are summarized in Table I.

Table I. Tomato β-galactosidase (TBG) cDNA sequence data. **Fiv full-length and 2 partial-length cDNAs were cloned and sequenced. The DNA and deduced amino acid sequence data is presented below**

CLONE	mRNA(kb)	kD	pI	N-LINK	TARGET
TBG1	3.2	90.8	6.2	2	ER/OUT
TBG2	3.0	97.0	6.2	6	PM
TBG3	2.8	90.5	8.2	1	ER/OUT
TBG4	2.6	77.9	8.9	3	OUT
TBG5	~3				
TBG6	~3				
TBG7	3.0	93.3	8.0	6	CHLOR

N-LINK = possible N-linked glycosylation sites; ER = endoplasmic reticulum; out = secreted; PM = tethered to plasma membrane; CHLOR = chloroplast

As indicated, nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment

For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Isolated nucleic acid molecules of the present invention include DNA molecules comprising an open reading frame (ORF) with an initiation codon at position 64 of the nucleotide sequence shown in Figure 2 [SEQ ID NO: 4]. Also included are DNA molecules comprising the coding sequence for the mature β-galactosidase II protein shown at positions 135-2532 of Figure 2 [SEQ ID NO: 4].

In addition, isolated nucleic acid molecules of the invention include DNA molecules which comprise a sequence substantially different from those described above but which, due to the degeneracy of the genetic code, still encode the β-galactosidase II protein. Of course, the genetic code and species-specific codon preferences are well known in the art. Thus, it would be routine for one skilled in the art to generate the degenerate variants described above, for instance, to optimize codon expression for a particular host (e.g., change codons in the plant mRNA to those preferred by a bacterial host such as *E. coli*). Preferably, this nucleic acid molecule will encode the mature polypeptide encoded by the above-described deposited cDNA clone.

The invention further provides an isolated nucleic acid molecule having the nucleotide sequence shown in Figure 2 [SEQ ID NO: 4] or a nucleic acid molecule having a sequence complementary to the above sequence. Such isolated molecules, particularly DNA molecules, are useful as probes for gene mapping, by *in situ* hybridization with chromosomes, and for detecting expression of the β-galactosidase II gene in plant tissue, for instance, by Northern blot analysis.

The present invention is further directed to nucleic acid molecules encoding portions of the nucleotide sequences described herein as well as to fragments of the isolated nucleic acid molecules described herein. In particular, the invention provides a polynucleotide having a nucleotide sequence representing the portion of Figure 2 [SEQ ID NO: 4] which consists of positions 1-2538 of Figure 2 [SEQ ID NO: 4].

In addition, the invention provides additional nucleic acid molecules having nucleotide sequences related to extensive portions of Figure 2 [SEQ ID NO: 4] which have been determined from the following related cDNA clones: TBG1-3 and TBG5-7 as shown in Figure 3, SEQ. NO's 1-3 and 5-7

In another aspect, the invention provides an isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a portion of the polynucleotide in a nucleic acid molecule of the invention described above, for instance, the cDNA clone shown in Figure 2 [SEQ ID NO: 4]. By "stringent hybridization conditions" is intended overnight incubation at 42° C in a solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml

denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65° C.

As indicated, nucleic acid molecules of the present invention which encode a β-galactosidase II polypeptide may include, but are not limited to those encoding the amino acid sequence of the mature polypeptide, by itself; and the coding sequence for the mature polypeptide and additional sequences, such as those encoding the about 1-23 amino acid leader sequence, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences.

Also discovered is an enhancer/promoter associated with expression of the genes encoding β-galactosidase. The inventors have characterized the expression profile of TBG2 mRNA and have cloned a lambda genomic cDNA. TBG2 is expressed before the onset of fruit ripening and continues at uniform level throught all the ripening stages. TBG2 has been found to be expressed in all fruit tissues and has also been found to be fruit specific. Experiments have shown TBG2 to be unaffected by ethylene. TBG2 is expressed in the ripening mutants rin, nor and Nr at the normal chronological time after anthesis. The promoter discovered would be useful to express any gene in the sense or antisense orientation, specifically in tomato fruit, in all tomato fruit tissues, starting before and continuing throughout the entire ripening process. The promoter could also be used to express any gene in the ripening mutants rin, nor and Nr without the need to gas the fruit with exogenous ethylene.

Variant and Mutant Polynucleotides

The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of the β-galactosidase II protein. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. *Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the β-galactosidase II protein or portions thereof. Also especially preferred in this regard are conservative substitutions.

Most highly preferred are nucleic acid molecules encoding the mature protein having the amino acid sequence shown in Figure 2 as pZBG2-1-4 or the mature β-galactosidase II amino acid sequence encoded by the deposited cDNA clone.

Further embodiments include an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 90%

identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical to a polynucleotide selected from the group consisting of: (a) a nucleotide sequence encoding the β-galactosidase II polypeptide having the complete amino acid sequence in Figure 2 [SEQ ID NO: 4] (b) a nucleotide sequence encoding the mature β-galactosidase II polypeptide shown in Figure 2 [SEQ ID NO: 4]; (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b) above.

Vectors and Host Cells

The present invention also relates to vectors which include the isolated DNA molecules of the present invention, host cells which are genetically engineered with the recombinant vectors, and the production of β-galactosidase II polypeptides or fragments thereof by recombinant techniques. The vector may be, for example, a phage, plasmid, viral or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli lac*, *trp*, *phoA* and *tac* promoters, the SV40 early and late promoters and promoters of

retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, StrepZBG2-1-4yces and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293 and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc., *supra*; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis *et al., Basic Methods In Molecular Biology* (1986).

Example

Tomato (*Lycopersicon esculentum* Mill., cv. 'Rutgers') plants were grown in a greenhouse using standard cultural practices. The ripening mutants, *ripening inhibitor* (*rin*), *non-ripening* (*nor*) and *never ripe* (*Nr*) (Tigchelaar *et al.*, 1978), were all in the 'Rutgers' background. Flowers were tagged at anthesis and fruit were harvested according to the number of days post-anthesis (dpa) or based on their surface color using ripeness stages as previously described (Mitcham *et al.*, 1989), the complete disclosure of which is hereby fully incorporated herein by reference. For gene expression studies, a variety of leaf, flower, and stem tissues were harvested from greenhouse-grown plants and roots were harvested from seedlings grown in basal tissue culture medium for 4 weeks after seed germination.

RNA Extraction

Fruits were processed immediately after harvest in the greenhouse by chilling on ice, excising the various tissues and freezing them in liquid nitrogen. Tissue samples were ground using a mortar and pestle and stored at -80°C. RNA was extracted using the method described in Verwoerd et al. (1989). Poly(A)RNA was purified from total RNA using oligo(dT) columns

(Pharmacia, Piscataway, NJ). RNA was quantified by measuring A_{260} using a dual beam spectrophotometer.

RT-PCR

Degenerate primers were designed based on the highest shared deduced amino acid sequence identity we found between an apple (accession number P48980), asparagus (P45582) and carnation (Q00662) β-galactosidase cDNA clones. The two primers used for the first reaction were BG5'E1 (WSNGGNWSNATHCAYTAYCC) and BG3'E (CCRTAYTCRTCNADNGGNGG). A second reaction was done on the products of the first reaction using BG5'I1 (ATHCARACNTAYGTNTTYTGG) and BG3'E. The degeneracy code for the primer sequences is N=a+t+c+g; H=a+t+c; B=t+c+g; D=a+t+g; V=a+c+g; R=a+g; Y=c+t; M=a+c; K=t+g; S=c+g; and W=a+t. The 5' and 3' primers corresponded to amino acids 72-78 and 321-315 of the apple clone, respectively. Amplification was done using AmpliTaq DNA polymerase (Perkin Elmer, Norwalk, CT) and standard PCR conditions using the cDNA made for the first cDNA library described below as a template (Ausubel *et al.*, 1987). PCR products were separated in an agarose gel and fragments of the expected size (approximately 750 bp) were purified, cloned into pCRscript (Stratagene, La Jolla, CA) , and sequenced.

cDNA library

Two cDNA libraries were constructed. The first comprised poly(A) RNA isolated from breaker, turning and pink fruit pericarp from 'Rutgers' plants.

The cDNA synthesis and library construction was done exactly according to the manufacturers instructions for the ZAP-cDNA Gigapack II Gold Cloning Kit (Stratagene), the complete disclosure of which is fully incorporated herein by reference. First-strand cDNA synthesis was primed using a poly(dT) primer and inserts were directionally cloned into the Uni-Zap XR vector using EcoRI and XhoI restriction sites. The second library comprised poly(A) RNA isolated from all fruit tissues (except seeds) from immature green, mature green, breaker, turning, pink, red-ripe and over-ripe fruit of 'Rutgers' plants. The cDNA synthesis and library construction was done exactly according to the manufacturers instructions for the SuperScript Lambda System for cDNA synthesis and • Cloning (GibcoBRL, Gaithersburg, MD). First-strand cDNA synthesis was primed using a oligo(dT) primer and cDNA inserts were directionally cloned into the • ZipLox cloning vector using SalI and NotI restriction sites. Both libraries were amplified and maintained using the host strains provided by the manufacturer, according to their instructions.

One of the clones (RT-PCR2-1) was used to screen 10^6 plaques from the tomato fruit cDNA libraries at low stringency (hybridization at 45°C, no formamide and final wash with 0.2X SSC at 42°C). Thirty positive cDNA clones were identified and partially sequenced. Complete sequencing and characterization of the RT-PCR and cDNA clones revealed the possibility of seven unique β-galactosidase genes.

DNA and RNA Gel Blot Analysis

Southern analysis was done using the 3' UTR of each full length clone and the RT-PCR clones as probes against restriction enzyme digested genomic DNA. DNA gel blot analysis was done essentially as described in Smith and Fedoroff (1995) except that 3 µg of genomic DNA was used for each digest. The genes corresponding to the clones appeared to be present as single copies (data not shown). The same probes were used to map 6 of the 7 genes using RFLPs of recombinant inbred lines and the loci names and map positions are shown in Table II (James Gioviannone, Texas A&M University, personal communication).

Table II. TBG loci map positions. Genes were maped by Southern analysis using RFLPs of recombinant inbred lines.

Gene	chromosome	map position
TBG1	12*	overlap of IL 12-2, IL 12-3
TBG2	9	IL 9-3
TBG3	3	IL 3-5
TBG4	12*	overlap of IL 12-2, IL 12-3
TBG5	11	IL 11-3
TBG6	2	overlap of IL 2-4, IL 2-5
TBG7	no RFLP	

*TBG1 and 4 are loosely linked

Total RNA (20 µg/ lane) was separated in a formaldehyde/Mops agarose gel, transferred to Hybond-N$^+$ nylon membrane (Amersham, Arlington Heights, IL), fixed by incubating for 2 h at 80°C, hybridized overnight in a

hybridization incubator (Robbins Scientific, Sunnyvale, CA) using a buffer described by Church and Gilbert (1984) washed to a final stringency of 0.1 X SSC with 0.2% SDS at 65°C, and autoradiographed essentially as described by Ausubel *et al.* (1987). An RNA ladder standard (GibcoBRL) was used to estimate the length of the RNAs. Probes were synthesized using a random priming kit with ^{32}P-dATP as the label (Boehringer Mannheim, Indianapolis, IN). Northern analysis was done using the 3' UTR of each full length clone and the RT-PCR clones as templates for probe synthesis. As a loading control, RNA blots were stripped and re-probed at a reduced hybridization and washing stringency using a soybean 26S rDNA fragment (Turano et al., 1997). For all hybridizations, ^{32}P(dATP)-labeled probe was diluted to 1-2 x 10^6 dpm/mL. The complete disclosures of the above references are fully incorporated herein by reference.

Sequence Analysis

Sequencing was done at the Iowa State University Sequencing Facility (Ames, IA) using a PCR-based dideoxynucleotide terminator protocol and an ABI automated sequencer (Applied Biosystems, Foster City, CA). The sequencing of both cDNA insert strands was done by primer walking. Nucleotide and deduced amino acid sequence comparisons against the databases were done using BLAST searches (Altschul *et al.*, 1990). Sequence data were analyzed and aligned using DNA Strider 1.2 (Marck, 1988) and MacDNAsis (Hitachi, San Bruno, CA) software. The complete disclosures of the above references are fully incorporated herein by reference.

Northern Blot Analysis

Tissue Specific Expression

Northern blot analysis was done to reveal which, if any, of the β-galactosidase genes had a fruit-specific expression pattern. With the exception of TBG2, transcripts of all clones were detected in non-fruit tissues (Fig. 4). Transcripts of TBG 1, 4, 5 and 6 were detected in all the tissues tested. TBG3 transcript was detected at low levels in root and stem tissues, while TBG7 transcript was detected in flower and stem tissues.

Temporal expression pattern in fruit

The temporal expression pattern of the seven genes in fruit tissue was examined using RNA extracted from all fruit tissues except seeds. Transcripts for all seven genes were detected during some stage of fruit development (Fig. 5). TBG1 and 3 had similar expression patterns and their transcripts were detected throughout the breaker to over-ripe stages. TBG2 had a unique expression pattern and its transcript was detected at a constant level from 30 dpp to the over ripe stage. TBG4 expression pattern was similar to TBG1 and 3, but differed in that the transcript level was significantly higher at the turning stage. TBG5 had a similar expression pattern to TBG4 during the ripening stages of development, however TBG5 transcript was also detected throughout all the earlier stages of fruit development. TBG6 had an interesting expression pattern and its transcript was only detected at high levels in all pre-ripening stages tested. TBG7 also had a unique expression pattern and its transcript was detected at very low levels throughout all the stages tested, and at moderate levels at 10 dpp and the over-ripe stage.

Spatial expression pattern in fruit

Northern blot analysis was also done to determine transcript accumulation in various fruit tissues. Since there were temporal differences in the expression patterns of the TBG genes both the mature green and turning fruit stages were used for RNA extractions (Fig. 6). Both TBG2 and TBG6 transcripts were detected in all mature green fruit tissues tested. TBG7 transcript was present in all fruit tissues tested except for locules. Both TBG1 and TBG4 transcripts were detected in RNA samples extracted from all turning stage fruit tissues. TBG4 transcript was notably more abundant in the peel. TBG3 and TBG5 expression patterns were unique and their transcripts were detected in all tissues except the outer pericarp and locular respectively.

Expression in normal versus mutant fruit

In order to better understand the potential roles of the TBG products and transcriptional regulatory mechanisms, northern analysis was performed using fruit tissue from the ripening mutants *rin, nor* and N^r. This analysis was important because it might give clues for preliminary determination of any potential ripening and/or softening role any of the TBGs might possess.

The results of mutant fruit Northern analysis suggested that the transcriptional regulation of TBG1, 2, 3, 5 and 7 was unaffected in mutant fruit tissue and that their transcripts were present in a normal chronological (dpp) pattern (Fig. 7). The abundance of TBG4 and 6 transcripts were however different in the mutant fruit. TBG4 transcript was not detected in fruit tissue of N^r and was detected at much lower levels in *rin* and *nor* than wild type fruit

tissues. Normally TBG6 transcripts are detectable at high levels throughout the early stages of fruit development but are not detectable after the mature green stage (40-42 dpp). TBG6 transcripts persisted even to 50 dpp in fruit of all three mutants.

Transcriptional regulation by ethylene

The northern analysis done using mutant and wild type fruit suggested that TBG4 expression might be up-regulated by ethylene and that TBG6 expression might be down-regulated by ethylene. In order to evaluate this hypothesis mature green fruit were harvested and subjected to a continuous flow of 10 ppm ethylene mixed in air. Control and ethylene-treated fruit were used for RNA extractions at 1, 2, 12 and 24 hours. The results of this experiment confirmed the findings from the mutant fruit northern analysis. As expected, the presence and abundance of TBG1, 2, 3, 5 and 7 transcripts was essentially unaffected in mature green tissues subjected to exogenous ethylene treatment (Fig. 8). However, TBG4 transcript abundance was increased in mature green tissues in the presence of ethylene. From the data presented it was unclear whether TBG6 transcript abundance was reduced by exogenous ethylene treatment since its transcript level was normally reduced at this stage of fruit development.

Enzyme activity

In order to determine the role of the TBG encoded products we initiated experiments to express the cDNA encoded enzymes using heterologous expression systems. Several E. coli expression systems were

tested, but the yield of product was very low due to toxicity (See the example below). Therefore we used a yeast expression system which secretes a mature amino-terminal-FLAG fusion protein into the culture medium. The TBG4 cDNA was tested first and resulted in the production of approximately 1 mg TBG4 active protein per 50 mls culture. TBG4 was used first because the cDNA codes for the enzyme β-galactosidase II which was purified from tomato fruit and has been characterized in some detail (Carey et al 1995, Smith et al 1998). Therefore we could compare the activity of the heterologous system-expressed protein to the native enzyme purified from tomato. The TBG4 protein was successfully affinity purified using an anti-FLAG affinity resin (Figure 9).

The affinity-purified TBG4 enzyme was shown to have β(1→4)-D-galactosidase activity by virtue of its ability to hydrolyze the synthetic substrate p-nitrophenyl-β-D-galactopyranoside (Smith et al. 1998). The enzyme can cleave galactosyl residues from a variety of cell wall substrates and therefore has exo-galactanase activity (Table III). The remaining full-length cDNA clones are currently being tested for successful expression of active enzyme. Preliminary results have shown that TBG1 codes for an enzyme which also has both β-D-galactosidase and exo-galactanase activity (Table III).

Table III. Cell wall degrading activity of TBG4 and TBG1 expressed in yeast. Removal of galactosyl residues from chelator soluble (CSP) and alkali soluble (ASP) pectin and hemicellulosic (HCF) cell wall fractions purified from tomato fruit.

		µg galactose released	
enzyme	substrate	boiled	live
[a]TBG4	CSP	0	5
	ASP	0	14.5
	HCF	0	4
[b]TBG1	ASP	0	1.2

2 mg substrate; 4 hours at 37°C
[a].005 units enzyme/rx
[b].0005 units enzyme/rx

pZBG2-1-4 Codes for a β-Galactosidase

The TBG4 ORF was cloned in-frame into the repressible/inducible bacterial expression vector pFLAG-CTC. The host strain XL1-Blue MR is a mutant strain containing no endogenous β-galactosidase activity nor α-complementation. Induction of gene transcription by (IPTG) caused the immediate cessation of *E. coli* growth at 30 to 37°C. However, induction at 20°C did allow for some limited *E. coli* growth. When clones containing the pZBG2-1-4 4 ORF were grown at 20°C and induced with IPTG, the cells slowly turned blue after 36 hrs growth in medium containing the β-galactosidase substrate X-GAL, (Figure 10). If not induced with IPTG, no blue color was seen, even after extended growth in media containing X-GAL. As an additional negative control, clones consisting of XL1-Blue MR transformed with the FLAG vector alone never showed any β-galactosidase activity with or without IPTG-induction, even after 7-days growth (Fig 10).

As a positive control for maximal β-galactosidase (derived from *E. coli* β-galactosidase) activity the cloning vector pGEM was transformed into the host strain DH5α and the results are also shown in Figure 10. Figure 10 shows the detection of β-galactosidase activity from pZBG2-1-4 expression in *E. coli.* Cells were harvested and extracts were prepared every 12 hours and the A_{615} measured. Cultures were grown with the addition of the chromogenic substrate X-GAL (open symbols) or X-GAL and the transcriptional inducer IPTG (closed symbols) in the medium. The vector used as a positive control for *E. coli* β-galactosidase activity was pGEM (■) and the vector used as a negative control and for expression was pFLAG-CTC either without (○,●) or containing the pZBG2-1-4 ORF (△,▲).

Effects on Plant Tissue Texture

To further demonstrate the function of TBG4 encoded β-galactosidase II the following experiments were carried out.

Fruit from tomato plants containing antisense constructs to suppress TBG4 mRNA were up to 40% firmer [compare means of parental line #1 with antisense line #2 in Figures 11A – 11E(1-4)] than fruit from the parental line. Among the transformants the line with the firmest fruit also had the lowest overall levels of TBG4 mRNA (Figure 12A,B). This correlation suggests that a reduction in TBG4 mRNA is associated with increased fruit firmness. Firmer fruit might result in (1) less shipping damage (a) less loss due to damage and (b) ability to harvest at later stage resulting in better flavor at market (2) longer

shelf life for both market and consumer. (3) better quality fruit for fresh slice market; fruit cut better at the pink/red stage when firmer.

Methods

To determine the function of TBG4 encoded β-galactosidase II, antisense constructs were made using the constitutively expressed 35S CaMV promoter to express TBG4 antisense RNA (Figure 13). Constructs were moved into tomato using Agrobacterium-mediated transformation. Four tomato cultivars have been transformed in order to evaluate the effect of TBG4 suppression on processing tomato (cv 'UC82b') fruit paste quality and three fresh pick cultivars. Of the fresh pick cultivars one is a soft fruit large cherry tomato (cv 'Ailsa Craig'), the second is a soft fruit old breeding line (cv 'Rutgers') and the third is a recently developed somewhat firm cultivar 'New Rutgers'. Among the lines where TBG4 mRNA is suppressed we expect to observe an increase in firmness and paste viscosity.

Texture

Although this project is nearly finished the complete biochemical and molecular analysis is not finished. The preliminary results on the analysis of the 'New Rutgers' cultivar is presented in Figures 11A – E(1-4) and 12A,B. In this example a fresh pick cultivar called 'New Rutgers' was used. Plants of the purchased seed were grown and allowed to self and the resulting seed was used as the parental control (line 1). Seven independent transformed plants (lines 2-8) containing TBG4 antisense constructs were grown and allowed to self. Transformation (T-DNA insertion) was confirmed by southern analysis

(data not shown). From each transformed line, five plants were grown along with 10 parental line plants. Fruit were tagged at the breaker stage (1st onset of color change) and were harvested at breaker plus 7 days. Data were taken using 15-20 fruit from each line. Each type of texture measurement was done twice for each fruit and fruit were subjected to 4 types of texture measurements using a Stable Micro System's TA-XT2i texture analyzer. The 4 measurements were; 1, 2-inch flat plate compression to 3 mm (Figure 1A), 2, 4 mm spherical indenter compression to 3 mm (Figure 11B), 3, 4 mm cylindrical indenter compression to 3 mm (Figure 11C) and 4, 4 mm cylindrical indenter puncture to 10 mm (Figure 11D). The summary of this data is shown in Figure 11E(1-4). In Figures 11A –E (1-4) line 1 was the parental line and lines 2-8 each represent an independent transformant containing one T-DNA copy of the TBG4 antisense construct. Statistical analysis (Duncans and Scheffé) of the data revealed that fruit from the transformed lines 3, 7 and 8 were not significantly different from the parental line but that transformed lines 2, 4, 5 and 6 were significantly firmer than the parental fruit. Most noteworthy is that fruit from transformed line 2 had fruit with a mean firmness that was 40% firmer than that of the parental line (Figures 11A-D).

Northern Blot Analysis

We are currently investigating any changes in the biochemical composition of fruit where TBG4 mRNA levels have been suppressed. These experiments are designed to show a link between increased fruit firmness and TBG4 mRNA suppression, TBG4 encoded enzyme activity suppression,

possible cell wall modification (e.g. increased galactosyl residue content) and a decrease in free galactose levels during fruit ripening.

These experiments are not complete, however some preliminary Northern blot experiments were done and the data is shown in Figure 12A,B. There is no parental or azygous control fruit RNA shown in Figure 12A,B because these plants were the last to grow and RNA extractions are just being done now. As a comparison of normal fruit TBG4 mRNA levels refer to Figure 5 above. The data from Figure 5 showed that TBG4 mRNA levels are low at the mature green stage, peak at the turning stage and are reduced at the red stage. All the lines except for 2 and 3 expressed antisense TBG4 mRNA (Figure 12A,B). The antisense transcripts appear as two bands, smaller in length than the endogenous mRNA. The two bands probably resulted from 1, the expected transcriptional stop signal provided by the NOS-terminator and 2, a cryptic transcriptional stop signal in the antisense TBG4 cDNA. The most notable result was in line 2 where no TBG4 mRNA was detected at the turning stage. Line 2 also had the firmest red fruit (see Figure 11A -D). The absence of detectable TBG4 mRNA probably was the result of cosupression of both the endogenous and antisense mRNAs. When compared to earlier blots (e.g. Figure 4), all of the lines appeared to have an overall reduced level of TBG4 mRNA, but it is impossible to assign numbers to this statement without the parental and azygous control RNA on the same Northern blot.

The specification discloses that β-galactosidase II polypeptide is involved in the degradation of cell wall pectin during fruit ripening. In the present invention, the role of β-galactosidases in tomato during fruit ripening and softening and the description of the cloning of a β-galactosidase cDNA clone

that codes for a β(1→4)galactan degrading enzyme, which is expressed in ripening tomato fruit tissues, has been shown.

The present work indicates that pZBG2-1-4 is a cDNA derived from the transcript of the TBG4 gene which codes for β-galactosidase II for the following reasons:

First, the deduced amino acid sequence of the highly conserved amino-terminal portion of the expected mature pZBG2-1-4 translation product matches almost exactly (28 of 30 amino acids) with the amino-terminal sequence of β-galactosidase II as purified by Carey *et al.* (1995) and designated TOMAA. Importantly, the two amino acids (KY) in the β-galactosidase II sequence (TOMAA), that do not match the pZBG2-1-4 deduced amino acid sequence of the present invention are believed to be incorrect since all plant β-galactosidase sequences in the database and four additional β-galactosidase-related cDNAs that were identified from tomato all match or have conserved substitutions with the deduced amino acid sequence of pZBG2-1-4 at these same two amino acid (ST) positions (Figure 3).

Second, the transcript detected by pZBG2-1-4 is present in normal ripening fruit at the same time that β-galactosidase II activity was detected (Figure 5; Carey *et al.*, 1995). Moreover, little or no transcript was detected in fruit at 45 and 50 dpa from the mutants *nor, rin* and *Nr* (Figure 7). This observation also coincides with the data presented by Carey *et al.* (1995) that β-galactosidase II activity remained at levels equal to mature green fruit and did not rise in fruit 45-65 dpa from *nor* or *rin* plants. Interestingly, Carrington and Pressey (1996) have reported that β-galactosidase II activity was only

detected in 'Rutgers' fruit after the turning stage of ripeness. The Northern data in the present invention indicates that maximum β-galactosidase II activity occurs only after the turning stage, assuming mRNA levels predict extractable enzyme activity (Figure 5).

Third, the apparent molecular weight of 77.9 kD and pI of 8.9 for the mature protein predicted from the pZBG2-1-4 sequence is similar to that determined for β-galactosidase II., Pressey (1983), estimated a molecular weight of 62 kD by gel-filtration column chromatography and a pI of 7.8 by isoelectric focusing, while Carey *et al.* (1995) estimated a molecular weight of 75 kD by SDS-PAGE and a pI of 9.8 by isoelectric focusing.

Fourth, enzyme produced from pZBG2-1-4 ORF using a heterologous yeast expression system has both β-galactosidase activity and exogalactinase activity.

LITERATURE CITED

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:**403-410

Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K, eds, (1987) Current Protocols in Molecular Biology. John Wiley and Sons, New York

Carey AT, Holt K, Picard S, Wilde R, Tucker GA, Bird CR, Schuch W, Seymour GB (1995) Tomato exo-(1→4)- β-D-galactanase. Isolation, changes during ripening in normal and mutant tomato fruit, and characterization of a related clone. Plant Physiol **108:**1099-1107

Carrington CM, Pressey R (1996) β-galactosidase II activity in relation to changes in cell wall galactosyl composition during tomato ripening. J Amer Soc Hort Sci **121:**132-136

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA **81:**1991-1995

DeVeau EJ, Gross KC, Huber DJ, Watada AE (1993) Degradation and solubilization of pectin by β-galactosidases purified from avocado mesocarp. Physiol Plant **87:**279-285

Fischer RL, Bennett AB (1991) Role of cell wall hydrolases in fruit ripening. Annu Rev Plant Physiol Plant Mol Bio **42:**675-703

Giovannoni JJ, DellaPenna D, Bennett AB, Fischer RL (1989) Expression of a chimeric polygalacturonase gene in transgenic *rin* (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. Plant Cell **1**:53-63

Golden KD, John MA, Kean EA (1993) β-Galactosidase from *Coffea arabica* and its role in fruit ripening. Phytochemistry **34:**355-360

Gross KC (1984) Fractionation and partial characterization of cell walls from normal and non-ripening mutant tomato fruit. Physiol Plant **62:**25-32

Gross KC, Sams CE (1984) Changes in cell wall neutral sugar composition during fruit ripening: A species survey. Phytochemistry **23:**2457-2461

Gross KC, Wallner SJ (1979) Degradation of cell wall polysaccharides during tomato fruit ripening. Plant Physiol **63:**117-121

Hall LN, Tucker GA, Smith CJS, Watson CF, Seymour GB, Bundick Y, Boniwell JM, Fletcher JD, Ray JA, Schuch W, Bird CR, Grierson D. (1993) Antisense inhibition of pectin esterase gene expression in transgenic tomatoes. Plant J **3**:121-129

Huber DJ (1983) The role of cell wall hydrolases in fruit softening. Hort Rev **5:**169-219

Kang IK, Suh SG, Gross KC, Byun JK (1994) *N*-terminal amino acid sequence of persimmon fruit β-galactosidase. Plant Physiol **105:** 975-979

Kim J, Gross KC, Solomos T (1991) Galactose metabolism and ethylene production during development and ripening of tomato fruit. Postharv Biol Technol **1:**67-80

Marck C (1988) DNA Strider: a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res **16:**1829-1836

Mitcham EJ, Gross KC, Ng TJ (1989) Tomato fruit cell wall synthesis during development and senescence. *In vivo* radiolabeling of cell wall fractions using [^{14}C]sucrose. Plant Physiol **89**:477-481

Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics **14**:897-911

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering **10**:1-6

Pressey R (1983) β-Galactosidases in ripening tomatoes. Plant Physiol **71**:132-135

Ross GS, Redgwell RJ, MacRae EA (1993) Kiwifruit β-galactosidase: isolation and activity against specific fruit cell-wall polysaccharides. Planta **189**:499-506

Ross GS, Wegrzyn T, MacRae EA, Redgwell RJ, (1994) Apple β-galactosidase. Activity against cell wall polysaccharides and characterization of a related cDNA clone. Plant Physiol **106**:521-528

Seymour GB, Gross KC (1996) Cell wall disassembly and fruit softening. Postharvest News Info **7**:45N-52N

Smith CJS, Watson CFS, Ray J, Bird CR, Morris PC, Schuch W, Grierson D (1988) Antisense RNA inhibition of polygalacturonase gene expression in transgenic tomatoes. Nature **334**:724-726

Smith DL, Fedoroff NV (1995) LRP1, a gene expressed in lateral and adventitious root primordia of Arabidopsis. Plant Cell **7**: 735-745

Smith DL, Starrett DA and Gross KC (1998) A gene coding for tomato fruit β-galatosidase II is expressed during fruit ripening. Plant Physiol. **117:** 417-423

Tieman DM, Harriman RW, Ramamohan G, Handa AK (1992) An antisense pectin methylesterase gene alters pectin chemistry and soluble solids in tomato fruit. Plant Cell **4:**667-679

Tigchelaar EC, McGlasson WB, Buescher RW (1978) Genetic regulation of tomato fruit ripening. HortScience **13:**508-513

Turano FJ, Thakkar SS, Fang T, Weisemann JM (1997) Characterization and expression of NAD(H) dependent glutamate dehydrogenase genes in Arabidopsis thaliana. Plant Physiol **113**: 1329-1341

Verwoerd TC, Dekker BMM, Hoekema A (1989) A small-scale procedure for the rapid isolation of plant RNAs. Nuc Acids Res **17:** 2362

Wegrzyn TF, MacRae EA (1992) Pectinesterase, polygalacturonase, and β-galactosidase during softening of ethylene-treated kiwifruit. HortScience **27:**900-902

What we claim is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the β-galactosidase II polypeptide having the complete amino acid sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO: 15 and SEQ ID NO: 16 and designated TBG1, TBG2, TBG3, TBG4, TBG5, TBG6 and TBG7, respectively as shown in Figure 2 or as encoded by the cDNA clone selected from the group consisting of cDNA clones contained in Gen Bank Accession No. AF023847, AF154420, AF154421, AF020390, AF154423, AF154424 and AF154422;

(b) a nucleotide sequence encoding the mature β-galactosidase II polypeptide having the amino acid sequence at about positions 24-724 selected from the group consisting of SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 and SEQ ID NO: 16 and designated TBG1, TBG2, TBG3, TBG4, TBG5, TBG6 and TBG7, respectively as shown in Figure 2 or as encoded by the cDNA clone selected from the group consisting of cDNA clones contained in Gen Bank Accession No. AF023847, AF154420, AF154421, AF020390, AF154423, AF154424 and AF154422; and

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b), above.

2. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 as shown in Figure 2.

3. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figure 2 (SEQ ID NO:4) encoding the β-galactosidase II polypeptide having the amino acid sequence designated TBG4 in Figure 2.

4. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figure 2 (SEQ ID NO:4) encoding the mature polypeptide having the amino acid sequence from about 24 to about 724 in the amino acid sequence designated TBG4 in Figure 2.

5. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in Gen Bank Accession No. AF023847.

6. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in Gen Bank Accession No. AF154420.

7. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in Gen Bank Accession No. AF154421.

8. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in Gen Bank Accession No. AF020390.

9. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in Gen Bank Accession No. AF154423.

10. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in Gen Bank Accession No. AF154424.

11. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in Gen Bank Accession No. AF154422.

12. An isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide having a nucleotide sequence identical to a nucleotide sequence in (a), (b), or (c) of claim 1 wherein said polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues.

13. An isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a β-galactosidase II polypeptide having an amino acid sequence in (a), (b), or (c) of claim 1.

14. A method for making a recombinant vector comprising inserting an isolated nucleic acid molecule of claim 1 into a vector.

15. A recombinant vector produced by the method of claim 14.

16. A method of making a recombinant host cell comprising introducing the recombinant vector of claim 15 into a host cell.

17. A recombinant host cell produced by the method of claim 16.

18. A recombinant method for producing β-galactosidase II polypeptide, comprising culturing the recombinant host cell of claim 17 under conditions such that said polypeptide is expressed and recovering said polypeptide.

19. An isolated β-galactosidase II polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

a) amino acid sequence at about positions 24-724 selected from the group consisting of sequences SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 and SEQ ID NO: 16 and designated TBG1, TBG2, TBG3, TBG4, TBG5, TBG6 and TBG7, respectively as shown in Figure 2; and

b) amino acid sequence as encoded by the cDNA clone selected from the group consisting of cDNA clones contained in Gen Bank Accession No. AF023847, AF154420, AF154421, AF020390, AF154423, AF154424 and AF154422.

20. An isolated polypeptide comprising an epitope-bearing portion of the β-galactosidase II protein.

21. An isolated antibody that binds specifically to a β-galactosidase II polypeptide of claim 20.

22. An isolated nucleic acid molecule nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the β-galactosidase II polypeptide having the complete amino acid sequence selected from the group consisting of SEQ ID NO:8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO: 15 and SEQ ID NO: 16 and designated TBG1, TBG2, TBG3, TBG4, TBG5, TBG6 and TBG7, respectively as shown in Figure 3 or as encoded by the cDNA clone selected from the group consisting of cDNA clones contained in Gen Bank Accession No. AF023847, AF154420, AF154421, AF020390, AF154423, AF154424 and AF154422;

(b) a nucleotide sequence encoding the mature β-galactosidase II polypeptide having the amino acid sequence at about positions 24-724 selected from the group consisting of sequences SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 and SEQ ID NO: 16 and designated TBG1, TBG2, TBG3, TBG4, TBG5, TBG6 and TBG7, respectively as shown in Figure 3 or as encoded by the cDNA clone selected from the group consisting of cDNA clones contained in Gen Bank Accession No. AF023847, AF154420, AF154421, AF020390, AF154423, AF154424 and AF154422; and

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b), above.

23. The nucleic acid molecule of claim 22 wherein said polynucleotide has a complete nucleotide sequence in Figure 2 selected from the group consisting of SEQ ID NOs: 1-3 and 5-7.

24. The nucleic acid molecule of claim 22 wherein said polynucleotide has a nucleotide sequence in Figure 2 selected from the group consisting of SEQ ID NOs: 1-3 and 5-7 encoding the β-galactosidase polypeptide having the complete amino acid sequence designated TBG1-3 and 5-7, respectively.

25. The nucleic acid molecule of claim 22 wherein said polynucleotide has the nucleotide sequence in Figure 2 selected from the group consisting of SEQ ID NOs: 1-3 and 5-7encoding the mature polypeptide having the amino acid sequence designated TBG1-3 and 5-7, respectively.

26. The nucleic acid molecule of claim 22 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in an Gen Bank Accession No. selected from the group consisting of ATCC Deposit No. selected from the group consisting of AF023847, AF154420, AF154421, AF020390, AF154423, AF154424 and AF154422.

27. A method of modifying cell wall metabolism in a plant which comprises transforming said plant with a DNA construct adapted to modify the activity of a β-galactosidase, growing said plant or its descendent and selecting a plant having modified cell wall characteristics, said construct comprising a transcriptional initiation region operative in plants operably linked to a DNA sequence encoding at least one β-galactosidase.

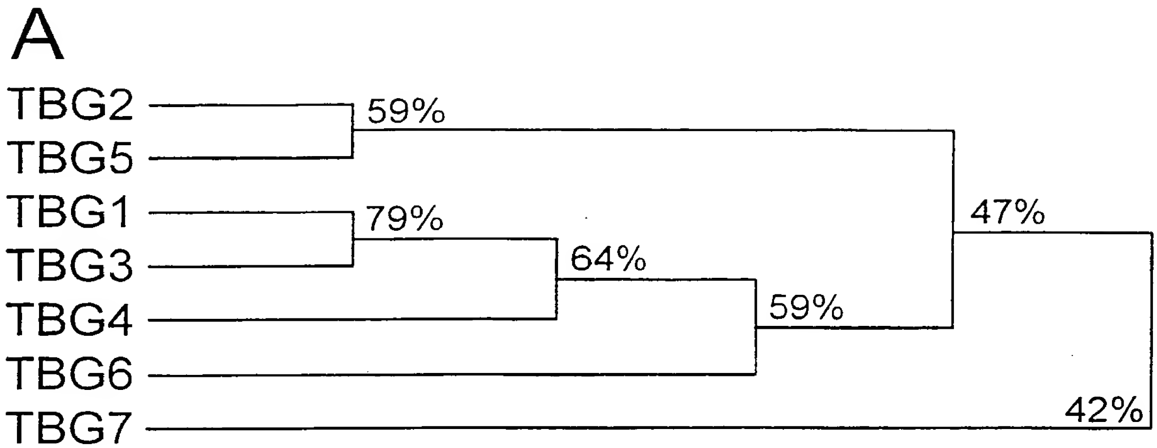
28. A method as claimed in claim 27, wherein said DNA sequence is selected from the group consisting of the sequences of nucleic acid molecules claimed in claim 1 or claim 22.

29. A plant cell transformed with a nucleic acid molecule as claimed in claim 1 or claim 22.

30. A plant derived from a plant cell as claimed in claim 29.

31. A plant seed derived from a plant as claimed in claim 30.

32. A method for modifying β-galactosidase gene expression in a plant comprising transforming said plant with a nucleic acid molecule as claimed in claim 1 or claim 22, growing the transformed plant and selecting a plant having modified β-galactosidase gene expression when compared with an untransformed plant.



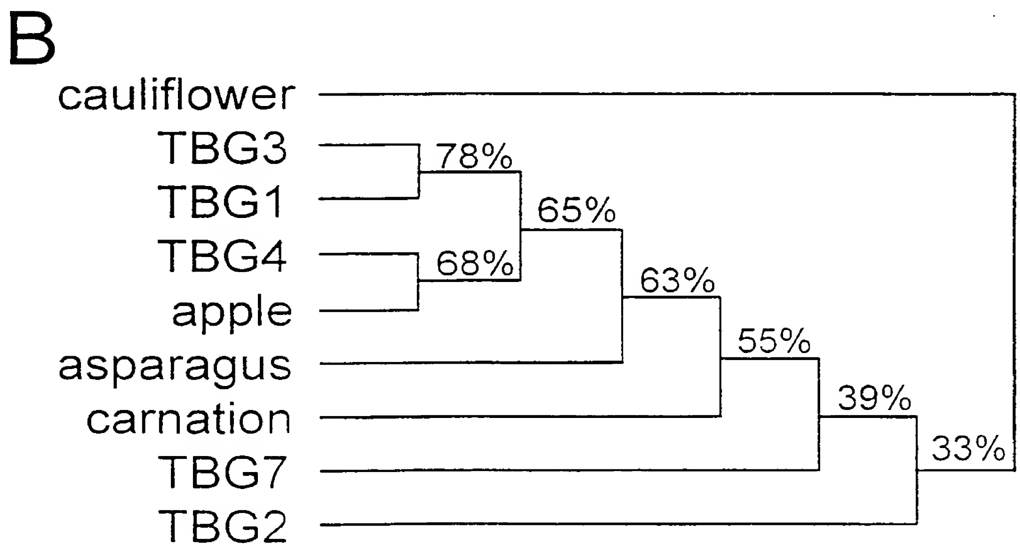


Figure 1. **β-Galactosidase phylogenetic tree based on shared amino acid sequence identity. A.** Tomato β-galactosidase (TBG) cDNAs. **B.** Plant β-galactosidases. Higgins-Sharp algorithm (UPGMA method)

Figure 2
Sh et 1 of 12

Gene/clone name: TBG1/pZBG2-1-10; accession number AF023847; Sequence ID number 1

TTTTTTCTTTGTTCTTTTGCTCAGCACTAG 30
31 AGCCTAGAAGAAGGAAAAAAAAGAAGTATGGACTAATGGAATAAACATAAAAAAGAGAGAAAAAAAAAAAAAGAAGAAAATTCTTCAGACAACA 122
123 AAAACAGCTGTTTTCCCTTCACTACTTTTTTTTTCCCAATCTCTATATAATTGCAAGAATAGAATAAAGTTTGCAACTTGATTAAAAAAAAA 214
215 GAATAATAAGCTGTGGGGGTAGGGAGGAAGTTAGTTCATTAGTTCATTGCCTTGTAAAGGCACAATCTTGATTCTTGATTTGTTGACAAAT 305

306 ATG GGT TTT TGG ATG GCA ATG TTG CTG ATG TTG TTA TTG TGT TTA TGG GTT TCT TGT GGA ATT GCT TCT 374
1 Met Gly Phe Trp Met Ala Met Leu Leu Met Leu Leu Leu Cys Leu Trp Val Ser Cys Gly Ile Ala Ser 23

375 GTT TCA TAT GAC CAT AAA GCT ATC ATT GTA AAT GGA CAA AGA AAA ATT CTC ATT TCT GGA TCC ATT CAC 443
24 Val Ser Tyr Asp His Lys Ala Ile Ile Val Asn Gly Gln Arg Lys Ile Leu Ile Ser Gly Ser Ile His 46

444 TAC CCT AGA AGC ACC CCT GAG ATG TGG CCA GAT CTT ATT CAG AAG GCA AAA GAA GGG GGA GTT GAT GTT 512
47 Tyr Pro Arg Ser Thr Pro Glu Met Trp Pro Asp Leu Ile Gln Lys Ala Lys Glu Gly Gly Val Asp Val 69

513 ATA CAG ACT TAT GTT TTC TGG AAT GGG CAT GAG CCT GAA GAA GGG AAA TAT TAT TTT GAA GAG AGG TAT 581
70 Ile Gln Thr Tyr Val Phe Trp Asn Gly His Glu Pro Glu Glu Gly Lys Tyr Tyr Phe Glu Glu Arg Tyr 92

582 GAT TTA GTG AAG TTC ATT AAA GTG GTG CAA GAA GCA GGA CTT TAT GTG CAT CTT AGG ATT GGA CCT TAT 650
93 Asp Leu Val Lys Phe Ile Lys Val Val Gln Glu Ala Gly Leu Tyr Val His Leu Arg Ile Gly Pro Tyr 115

651 GCA TGT GCT GAA TGG AAT TTT GGG GGT TTT CCT GTT TGG CTG AAG TAT GTT CCA GGT ATT AGT TTC AGA 719
116 Ala Cys Ala Glu Trp Asn Phe Gly Gly Phe Pro Val Trp Leu Lys Tyr Val Pro Gly Ile Ser Phe Arg 138

720 ACA AAC AAT GAG CCA TTC AAG GCT GCA ATG CAA AAG TTC ACT ACT AAG ATT GTT GAT ATG ATG AAA GCA 788
139 Thr Asn Asn Glu Pro Phe Lys Ala Ala Met Gln Lys Phe Thr Thr Lys Ile Val Asp Met Met Lys Ala 161

789 GAA AAG CTC TAT GAA ACT CAG GGT GGT CCA ATT ATT CTA TCT CAG ATA GAA AAT GAA TAT GGA CCT ATG 857
162 Glu Lys Leu Tyr Glu Thr Gln Gly Gly Pro Ile Ile Leu Ser Gln Ile Glu Asn Glu Tyr Gly Pro Met 184

858 GAG TGG GAA CTA GGT GAA CCT GGT AAA GTT TAC TCA GAA TGG GCA GCC AAA ATG GCT GTG GAT CTT GGC 926
185 Glu Trp Glu Leu Gly Glu Pro Gly Lys Val Tyr Ser Glu Trp Ala Ala Lys Met Ala Val Asp Leu Gly 207

927 ACT GGT GTC CCA TGG ATC ATG TGC AAG CAA GAT GAT GTC CCT GAT CCT ATT ATT AAT ACT TGC AAT GGT 995
208 Thr Gly Val Pro Trp Ile Met Cys Lys Gln Asp Asp Val Pro Asp Pro Ile Ile Asn Thr Cys Asn Gly 230

996 TTC TAC TGT GAC TAC TTC ACA CCA AAT AAG GCT AAT AAA CCC AAG ATG TGG ACT GAA GCC TGG ACA GCC 1064
231 Phe Tyr Cys Asp Tyr Phe Thr Pro Asn Lys Ala Asn Lys Pro Lys Met Trp Thr Glu Ala Trp Thr Ala 253

1065 TGG TTT ACC GAA TTT GGA GGT CCA GTT CCT TAC CGT CCT GCA GAG GAT ATG GCA TTT GCT GTC GCA AGA 1133
254 Trp Phe Thr Glu Phe Gly Gly Pro Val Pro Tyr Arg Pro Ala Glu Asp Met Ala Phe Ala Val Ala Arg 276

1134 TTT ATA CAA ACG GGA GGC TCC TTC ATC AAT TAC TAT ATG TAT CAT GGA GGA ACA AAC TTT GGA AGG ACT 1202
277 Phe Ile Gln Thr Gly Gly Ser Phe Ile Asn Tyr Tyr Met Tyr His Gly Gly Thr Asn Phe Gly Arg Thr 299

1203 TCT GGT GGC CCA TTT ATT GCT ACT AGT TAT GAT TAT GAT GCA CCC CTA GAT GAA TTT GGG TCA TTA CGG 1271
300 Ser Gly Gly Pro Phe Ile Ala Thr Ser Tyr Asp Tyr Asp Ala Pro Leu Asp Glu Phe Gly Ser Leu Arg 322

1272 CAG CCT AAA TGG GGT CAT CTG AAA GAT CTA CAT AGA GCA ATA AAG CTC TGT GAG CCA GCT TTA GTA TCT 1340
323 Gln Pro Lys Trp Gly His Leu Lys Asp Leu His Arg Ala Ile Lys Leu Cys Glu Pro Ala Leu Val Ser 345

1341 GTA GAT CCA ACT GTG ACA TCC TTA GGA AAC TAT CAA GAG GCA CGT GTT TTC AAG TCA GAG TCT GGG GCC 1409
346 Val Asp Pro Thr Val Thr Ser Leu Gly Asn Tyr Gln Glu Ala Arg Val Phe Lys Ser Glu Ser Gly Ala 368

1410 TGC GCT GCC TTC CTA GCA AAT TAC AAC CAG CAC TCT TTT GCT AAA GTG GCA TTT GGG AAC ATG CAT TAT 1478
369 Cys Ala Ala Phe Leu Ala Asn Tyr Asn Gln His Ser Phe Ala Lys Val Ala Phe Gly Asn Met His Tyr 391

1479 AAC TTG CCA CCC TGG TCT ATC AGC ATT CTT CCC GAC TGC AAG AAC ACT GTC TAT AAT ACT GCA AGG GTT 1547
392 Asn Leu Pro Pro Trp Ser Ile Ser Ile Leu Pro Asp Cys Lys Asn Thr Val Tyr Asn Thr Ala Arg Val 414

1548 GGT GCT CAA AGT GCT CAG ATG AAG ATG ACT CCA GTC AGT AGA GGA TTC TCA TGG GAG TCA TTC AAT GAA 1616
415 Gly Ala Gln Ser Ala Gln Met Lys Met Thr Pro Val Ser Arg Gly Phe Ser Trp Glu Ser Phe Asn Glu 437

Figure 2
Sh et 2 of 12

Gene/clone name: TBG1/pZBG2[redacted]10; accession number AF023847; Sequ[redacted]e ID number 1 cont.

```
1617 GAC GCA GCA TCG CAT GAA GAC GAC ACT TTC ACA GTT GTT GGG TTA TTG GAG CAG ATT AAT ATC ACA AGA   1685
 438 Asp Ala Ala Ser His Glu Asp Asp Thr Phe Thr Val Val Gly Leu Leu Glu Gln Ile Asn Ile Thr Arg    460

1686 GAT GTA TCT GAT TAC TTG TGG TAT ATG ACT GAC ATT GAG ATT GAT CCA ACA GAA GGA TTT TTG AAT AGT   1754
 461 Asp Val Ser Asp Tyr Leu Trp Tyr Met Thr Asp Ile Glu Ile Asp Pro Thr Glu Gly Phe Leu Asn Ser    483

1755 GGA AAT TGG CCT TGG CTT ACT GTC TTT TCT GCT GGC CAT GCA TTG CAT GTA TTC GTG AAT GGT CAA TTA   1823
 484 Gly Asn Trp Pro Trp Leu Thr Val Phe Ser Ala Gly His Ala Leu His Val Phe Val Asn Gly Gln Leu    506

1824 GCA GGA ACT GTG TAC GGA AGT TTA GAA AAC CCA AAA CTA ACT TTC AGC AAC GGT ATA AAT CTG AGA GCT   1892
 507 Ala Gly Thr Val Tyr Gly Ser Leu Glu Asn Pro Lys Leu Thr Phe Ser Asn Gly Ile Asn Leu Arg Ala    529

1893 GGT GTG AAC AAG ATT TCT CTG CTA AGC ATT GCT GTT GGT CTT CCG AAC GTT GGC CCT CAT TTT GAG ACA   1961
 530 Gly Val Asn Lys Ile Ser Leu Leu Ser Ile Ala Val Gly Leu Pro Asn Val Gly Pro His Phe Glu Thr    552

1962 TGG AAT GCT GGT GTT CTT GGA CCA GTT TCA CTT AAT GGA CTT AAT GAA GGA ACA AGA GAT TTA ACA TGG   2030
 553 Trp Asn Ala Gly Val Leu Gly Pro Val Ser Leu Asn Gly Leu Asn Glu Gly Thr Arg Asp Leu Thr Trp    575

2031 CAG AAA TGG TTC TAC AAG GTT GGT CTA AAA GGA GAA GCC CTG AGT CTT CAT TCA CTC AGT GGT AGC CCA   2099
 576 Gln Lys Trp Phe Tyr Lys Val Gly Leu Lys Gly Glu Ala Leu Ser Leu His Ser Leu Ser Gly Ser Pro    598

2100 TCC GTG GAG TGG GTG GAA GGC TCT TTA GTG GCT CAG AAG CAG CCA CTC AGT TGG TAT AAG ACT ACA TTC   2168
 599 Ser Val Glu Trp Val Glu Gly Ser Leu Val Ala Gln Lys Gln Pro Leu Ser Trp Tyr Lys Thr Thr Phe    621

2169 AAT GCT CCA GAT GGA AAT GAA CCT TTG GCT TTA GAT ATG AAT ACC ATG GGC AAA GGT CAA GTA TGG ATA   2237
 622 Asn Ala Pro Asp Gly Asn Glu Pro Leu Ala Leu Asp Met Asn Thr Met Gly Lys Gly Gln Val Trp Ile    644

2238 AAT GGT CAG AGC CTC GGA CGC CAC TGG CCT GCA TAT AAA TCA TCT GGA AGT TGT AGT GTC TGT AAC TAT   2306
 645 Asn Gly Gln Ser Leu Gly Arg His Trp Pro Ala Tyr Lys Ser Ser Gly Ser Cys Ser Val Cys Asn Tyr    667

2307 ACT GGC TGG TTT GAT GAG AAA AAG TGC CTA ACT AAC TGT GGT GAG GGC TCA CAA AGA TGG TAC CAC GTA   2375
 668 Thr Gly Trp Phe Asp Glu Lys Lys Cys Leu Thr Asn Cys Gly Glu Gly Ser Gln Arg Trp Tyr His Val    690

2376 CCC CGG TCT TGG CTG TAT CCT ACT GGA AAT TTG TTA GTT GTA TTC GAG GAA TGG GGA GGA GAT CCT TAT   2444
 691 Pro Arg Ser Trp Leu Tyr Pro Thr Gly Asn Leu Leu Val Val Phe Glu Glu Trp Gly Gly Asp Pro Tyr    713

2445 GGA ATC ACT TTA GTC AAA AGA GAA ATA GGG AGT GTT TGT GCT GAT ATA TAT GAG TGG CAA CCA CAG TTA   2513
 714 Gly Ile Thr Leu Val Lys Arg Glu Ile Gly Ser Val Cys Ala Asp Ile Tyr Glu Trp Gln Pro Gln Leu    736

2514 TTG AAT TGG CAG AGG CTA GTA TCT GGT AAG TTT GAC AGA CCT CTC AGA CCT AAA GCC CAT CTT AAG TGT   2582
 737 Leu Asn Trp Gln Arg Leu Val Ser Gly Lys Phe Asp Arg Pro Leu Arg Pro Lys Ala His Leu Lys Cys    759

2583 GCA CCT GGT CAG AAG ATT TCT TCA ATC AAA TTT GCA AGC TTT GGA ACA CCA GAG GGA GTT TGT GGG AAC   2651
 760 Ala Pro Gly Gln Lys Ile Ser Ser Ile Lys Phe Ala Ser Phe Gly Thr Pro Glu Gly Val Cys Gly Asn    782

2652 TTC CAG CAG GGA AGC TGC CAT GCT CCG CGC TCA TAT GAT GCT TTC AAA AAG AAT TGT GTT GGG AAA GAG   2720
 783 Phe Gln Gln Gly Ser Cys His Ala Pro Arg Ser Tyr Asp Ala Phe Lys Lys Asn Cys Val Gly Lys Glu    805

2721 TCT TGC TCA GTA CAG GTA ACA CCA GAG AAT TTT GGA GGT GAT CCA TGT CGA AAC GTT CTA AAG AAA CTC   2789
 806 Ser Cys Ser Val Gln Val Thr Pro Glu Asn Phe Gly Gly Asp Pro Cys Arg Asn Val Leu Lys Lys Leu    828

2790 TCA GTG GAA GCC ATT TGT AGT TGA TGATTCTGAGTATACAAGTGAAAAAATACTTGAACCACTCATATAAACATTTTTCAAACG   2873
 829 Ser Val Glu Ala Ile Cys Ser ***                                                                 836

2874 AGCTACTAGACATCCATTAACCCACACTACCATTTTTTGGCTTTGCTGGGGTTGAAGTTGTACAGTTAAGCAACACACCTCTTTGATCAAAG   2965
2966 CTCACCTGATTATGAAGATGATTGACGAAAGATTCTGTACATGTAAGGTTTCGTCTAATTACACATACAGATATGATTCTTGATGAATCGAT   3057
3058 GTGCAAATTTTGTTTGTGTTAGGGTGAGAGAGACTTGAAAAGCATTTTGCTTTCATGATGTTCTACATTATACAATCATAATGTAAGTAAGC   3149
3150 AAGCAATAATTCATTGCTTTGCACATTGAAAAATGCATTTTACTATGTTGCAGTACAAAAAAAAAAAAAAAAAAA                     3224
```

Figure 2
Sheet 3 of 12

Gene/clone name: TBG2/pZBG2-1-12; accession number AF154420; Sequence ID number 2

```
1                                                                                              GG     2

   3 AGC AGA AGA AAA ACA CTG AAT TTT CCG TTA ATA CTA ACG GTG TTA ACT ATC CAC TTT GTG ATC GTC GCC   71
   1 Ser Arg Arg Lys Thr Leu Asn Phe Pro Leu Ile Leu Thr Val Leu Thr Ile His Phe Val Ile Val Ala   23

  72 GGC GAG TAT TTC AAG CCG TTC AAT GTC ACC TAC GAT AAC CGA GCT CTC ATC ATC GGC GGT AAA CGC CGT  140
  24 Gly Glu Tyr Phe Lys Pro Phe Asn Val Thr Tyr Asp Asn Arg Ala Leu Ile Ile Gly Gly Lys Arg Arg   46

 141 ATG CTT ATC TCC GCC GGA ATT CAC TAC CCT CGC GCC ACT CCT GAG ATG TGG CCC ACA TTG ATA GCT AGG  209
  47 Met Leu Ile Ser Ala Gly Ile His Tyr Pro Arg Ala Thr Pro Glu Met Trp Pro Thr Leu Ile Ala Arg   69

 210 AGC AAA GAA GGT GGT GCA GAT GTC ATC GAG ACT TAT ACA TTT TGG AAT GGT CAT GAG CCA ACC AGG GGA  278
  70 Ser Lys Glu Gly Gly Ala Asp Val Ile Glu Thr Tyr Thr Phe Trp Asn Gly His Glu Pro Thr Arg Gly   92

 279 CAG TAC AAT TTT GAA GGA AGA TAT GAT ATT GTC AAG TTC GCA AAG CTA GTC GGA TCT CAT GGA CTG TTC  347
  93 Gln Tyr Asn Phe Glu Gly Arg Tyr Asp Ile Val Lys Phe Ala Lys Leu Val Gly Ser His Gly Leu Phe  115

 348 CTC TTT ATT CGA ATA GGT CCT TAT GCC TGT GCA GAA TGG AAC TTC GGG GGA TTC CCC ATA TGG CTT CGT  416
 116 Leu Phe Ile Arg Ile Gly Pro Tyr Ala Cys Ala Glu Trp Asn Phe Gly Gly Phe Pro Ile Trp Leu Arg  138

 417 GAT ATA CCT GGA ATA GAA TTT CGA ACA GAT AAT GCA CCA TTC AAG GAG GAG ATG GAG CGC TAT GTT AAA  485
 139 Asp Ile Pro Gly Ile Glu Phe Arg Thr Asp Asn Ala Pro Phe Lys Glu Glu Met Glu Arg Tyr Val Lys  161

 486 AAG ATA GTT GAT CTT ATG ATA TCT GAG TCG CTC TTT TCG TGG CAA GGT GGT CCT ATC ATT TTG CTG CAG  554
 162 Lys Ile Val Asp Leu Met Ile Ser Glu Ser Leu Phe Ser Trp Gln Gly Gly Pro Ile Ile Leu Leu Gln  184

 555 ATT GAA AAT GAA TAT GGA AAT GTT GAA AGC TCA TTC GGT CCC AAG GGG AAG TTA TAT ATG AAA TGG GCT  623
 185 Ile Glu Asn Glu Tyr Gly Asn Val Glu Ser Ser Phe Gly Pro Lys Gly Lys Leu Tyr Met Lys Trp Ala  207

 624 GCT GAA ATG GCT GTT GGT CTT GGT GCT GGT GTT CCA TGG GTC ATG TGC AGG CAA ACT GAT GCT CCA GAA  692
 208 Ala Glu Met Ala Val Gly Leu Gly Ala Gly Val Pro Trp Val Met Cys Arg Gln Thr Asp Ala Pro Glu  230

 693 TAC ATC ATA GAT ACT TGT AAT GCA TAC TAT TGT GAT GGG TTC ACG CCG AAT TCC GAG AAG AAA CCG AAA  761
 231 Tyr Ile Ile Asp Thr Cys Asn Ala Tyr Tyr Cys Asp Gly Phe Thr Pro Asn Ser Glu Lys Lys Pro Lys  253

 762 ATT TGG ACT GAG AAT TGG AAT GGA TGG TTT GCA GAT TGG GGT GAA AGA CTT CCA TAT AGA CCT TCC GAG  830
 254 Ile Trp Thr Glu Asn Trp Asn Gly Trp Phe Ala Asp Trp Gly Glu Arg Leu Pro Tyr Arg Pro Ser Glu  276

 831 GAT ATT GCA TTT GCA ATT GCT CGT TTC TTT CAA CGT GGG GGC AGC TTA CAG AAC TAT TAT ATG TAT TTT  899
 277 Asp Ile Ala Phe Ala Ile Ala Arg Phe Phe Gln Arg Gly Gly Ser Leu Gln Asn Tyr Tyr Met Tyr Phe  299

 900 GGT GGG ACA AAT TTT GGC CGG ACT GCT GGT GGC CCA ACT CAA ATC ACT AGC TAT GAT TAT GAT GCT CCA  968
 300 Gly Gly Thr Asn Phe Gly Arg Thr Ala Gly Gly Pro Thr Gln Ile Thr Ser Tyr Asp Tyr Asp Ala Pro  322

 969 CTG GAT GAA TAT GGA CTA CTA CGT CAA CCT AAA TGG GGC CAT TTG AAG GAT CTG CAT GCT GCT ATA AAG 1037
 323 Leu Asp Glu Tyr Gly Leu Leu Arg Gln Pro Lys Trp Gly His Leu Lys Asp Leu His Ala Ala Ile Lys  345

1038 CTT TGT GAA CCA GCT CTT GTT GCT GCT GAT TCA CCT CAG TAT ATT AAA CTG GGA CCA AAA CAG GAG GCA 1106
 346 Leu Cys Glu Pro Ala Leu Val Ala Ala Asp Ser Pro Gln Tyr Ile Lys Leu Gly Pro Lys Gln Glu Ala  368

1107 CAT GTC TAT CGT GGA ACA TCC AAC AAC ATT GGC CAA TAT ATG TCC TTA AAT GAA GGC ATA TGC GCA GCA 1175
 369 His Val Tyr Arg Gly Thr Ser Asn Asn Ile Gly Gln Tyr Met Ser Leu Asn Glu Gly Ile Cys Ala Ala  391

1176 TTT ATT GCA AAT ATT GAT GAA CAT GAA TCA GCA ACA GTG AAA TTT TAC GGT CAA GAG TTC ACT TTA CCT 1244
 392 Phe Ile Ala Asn Ile Asp Glu His Glu Ser Ala Thr Val Lys Phe Tyr Gly Gln Glu Phe Thr Leu Pro  414

1245 CCA TGG TCA GTG GTA TTC TGC CAG ATT GCA GAA ATA CAG CTT TCA ACA CAG CTA AGG TGG GGG CAC AAA 1313
 415 Pro Trp Ser Val Val Phe Cys Gln Ile Ala Glu Ile Gln Leu Ser Thr Gln Leu Arg Trp Gly His Lys  437

1314 CTT CAA TCA AAA CAG TGG GCT CAG ATT CTG TTT CAG TTG GGA ATA ATT CTT TGT TTC TAC AAG TTA TCA 1382
 438 Leu Gln Ser Lys Gln Trp Ala Gln Ile Leu Phe Gln Leu Gly Ile Ile Leu Cys Phe Tyr Lys Leu Ser  460
```

Figure 2
Sheet 4 of 12
Gene/clone name: TBG2/pZBG2-[redacted]2; accession number AF154420; Sequ[redacted] ID number 2 cont.

```
1383 CTA AAA GCA AGC TCG GAA AGT TTT TCA CAA TCT TGG ATG ACA TTG AAG GAG CCA CTT GGT GTG TGG GGT  1451
 461 Leu Lys Ala Ser Ser Glu Ser Phe Ser Gln Ser Trp Met Thr Leu Lys Glu Pro Leu Gly Val Trp Gly   483

1452 GAC AAG AAT TTC ACT TCT AAA GGA ATA CTG GAG CAT CTG AAT GTG ACA AAA GAC CAG TCT GAT TAC CTG  1520
 484 Asp Lys Asn Phe Thr Ser Lys Gly Ile Leu Glu His Leu Asn Val Thr Lys Asp Gln Ser Asp Tyr Leu   506

1521 TGG TAT CTG ACC AGG ATA TAT ATT TCT GAT GAT GAC ATC TCA TTT TGG GAG GAA AAT GAT GTT AGT CCA  1589
 507 Trp Tyr Leu Thr Arg Ile Tyr Ile Ser Asp Asp Asp Ile Ser Phe Trp Glu Glu Asn Asp Val Ser Pro   529

1590 ACA ATT GAT ATT GAT AGC ATG CGT GAT TTT GTT CGC ATT TTT GTT AAT GGG CAG CTT GCA GGT AGT GTG  1658
 530 Thr Ile Asp Ile Asp Ser Met Arg Asp Phe Val Arg Ile Phe Val Asn Gly Gln Leu Ala Gly Ser Val   552

1659 AAA GGC AAA TGG ATC AAG GTG GTT CAA CCT GTT AAG CTG GTT CAG GGA TAC AAC GAC ATA CTG CTA TTA  1727
 553 Lys Gly Lys Trp Ile Lys Val Val Gln Pro Val Lys Leu Val Gln Gly Tyr Asn Asp Ile Leu Leu Leu   575

1728 TCT GAG ACG GTG GGA TTG CAG AAT TAT GGT GCC TTC TTG GAG AAG GAT GGG GCA GGT TTT AAA GGT CAG  1796
 576 Ser Glu Thr Val Gly Leu Gln Asn Tyr Gly Ala Phe Leu Glu Lys Asp Gly Ala Gly Phe Lys Gly Gln   598

1797 ATA AAG CTT ACA GGA TGC AAA AGC GGG GAT ATC AAT CTC ACA ACA TCT TTA TGG ACC TAC CAG GTG GGG  1865
 599 Ile Lys Leu Thr Gly Cys Lys Ser Gly Asp Ile Asn Leu Thr Thr Ser Leu Trp Thr Tyr Gln Val Gly   621

1866 CTT AGA GGC GAA TTC CTG GAA GTA TAT GAT GTC AAT AGT ACT GAA AGT GCA GGA TGG ACT GAG TTT CCC  1934
 622 Leu Arg Gly Glu Phe Leu Glu Val Tyr Asp Val Asn Ser Thr Glu Ser Ala Gly Trp Thr Glu Phe Pro   644

1935 ACT GGT ACA ACT CCG TCA GTC TTT TCG TGG TAC AAG ACA AAG TTT GAT GCC CCA GGC GGG ACA GAT CCA  2003
 645 Thr Gly Thr Thr Pro Ser Val Phe Ser Trp Tyr Lys Thr Lys Phe Asp Ala Pro Gly Gly Thr Asp Pro   667

2004 GTT GCT CTT GAT TTT AGT AGC ATG GGA AAA GGT CAG GCA TGG GTT AAT GGC CAC CAT GTA GGA AGA TAT  2072
 668 Val Ala Leu Asp Phe Ser Ser Met Gly Lys Gly Gln Ala Trp Val Asn Gly His His Val Gly Arg Tyr   690

2073 TGG ACT TTG GTT GCA CCA AAT AAT GGA TGT GGA AGA ACT TGT GAT TAT CGT GGT GCT TAC CAC TCT GAT  2141
 691 Trp Thr Leu Val Ala Pro Asn Asn Gly Cys Gly Arg Thr Cys Asp Tyr Arg Gly Ala Tyr His Ser Asp   713

2142 AAA TGT AGG ACA AAC TGT GGA GAG ATT ACT CAG GCC TGG TAC CAT ATA CCT AGA TCA TGG CTA AAG ACA  2210
 714 Lys Cys Arg Thr Asn Cys Gly Glu Ile Thr Gln Ala Trp Tyr His Ile Pro Arg Ser Trp Leu Lys Thr   736

2211 TTA AAT AAT GTA CTA GTT ATC TTT GAA GAA ACA GAT AAA ACT CCG TTT GAT ATT TCC ATT TCT ACG CGT  2279
 737 Leu Asn Asn Val Leu Val Ile Phe Glu Glu Thr Asp Lys Thr Pro Phe Asp Ile Ser Ile Ser Thr Arg   759

2280 TCT ACT GAA ACC ATT TGT GCT CAA GTA TCG GAA AAG CAC TAT CCA CCT CTA CAT AAG TGG TCT CAT TCG  2348
 760 Ser Thr Glu Thr Ile Cys Ala Gln Val Ser Glu Lys His Tyr Pro Pro Leu His Lys Trp Ser His Ser   782

2349 GAG TTT GAC AGA AAG TTG TCT CTG ATG GAT AAA ACA CCA GAA ATG CAC TTG CAG TGT GAC GAA GGA CAT  2417
 783 Glu Phe Asp Arg Lys Leu Ser Leu Met Asp Lys Thr Pro Glu Met His Leu Gln Cys Asp Glu Gly His   805

2418 ACA ATC TCT TCT ATT GAA TTT GCA AGC TAT GGA AGT CCG AAT GGC AGC TGT CAA AAG TTC TCA CAA GGA  2486
 806 Thr Ile Ser Ser Ile Glu Phe Ala Ser Tyr Gly Ser Pro Asn Gly Ser Cys Gln Lys Phe Ser Gln Gly   828

2487 AAA TGC CAT GCT GCA AAT TCC TTG TCT GTT GTA TCT CAG GCT TGT ATA GGA AGA ACT AGT TGC AGC ATT  2555
 829 Lys Cys His Ala Ala Asn Ser Leu Ser Val Val Ser Gln Ala Cys Ile Gly Arg Thr Ser Cys Ser Ile   851

2556 GGC ATT TCC AAT GGT GTA TTT GGA GAT CCA TGT CGA CAC GTT GTG AAG AGT TTG GCT GTT CAA GCA AAA  2624
 852 Gly Ile Ser Asn Gly Val Phe Gly Asp Pro Cys Arg His Val Val Lys Ser Leu Ala Val Gln Ala Lys   874

2625 TGC TCA CCA CCA CCA GAC CTC AGC ACT TCA GCT TCC TCG TGA GGAGACTCTGGTAACACGTTAACCTTTTAGAACGAA  2702
 875 Cys Ser Pro Pro Pro Asp Leu Ser Thr Ser Ala Ser Ser ***                                       888

2703 ACGATCCCTTAAAGTCCACTCGTTCCCCTGCCCCCGAGCCCTCTGCTACATTTCTCAGATCGCATCGTTACAATCAGGCGGAGAAAACGTAC  2794
2795 ATGGACGATTTTACTTGTAAATATTTGGTTACTGTATATAAAATGAAAGGAATAATGTTGCTTATGCATATGAGCTGCAAATTATATGACAA  2886
2887 AGTAACAAATGAAAATAGAAAACTCCTGTCTGTCAAAGAATTTTAACAACACCATTTATTAAAAGTTAGTTAACATGATTAAAAAAAAAAAA  2978
2979 AAAAAA                                                                                          2984
```

Figure 2
Sheet 5 of 12

Gene/clone name: TBG3/p2- 0c/bl; accession number AF154421; Sequence ID number 3

AGAGTTCATTATTTTTTTTGCATTTTGAAA 30
1
31 AAGAGGAAAAAAATAAAGTTAAAGGGGGGGGAAAAAGTTTTCATTTTGCCTTAAAAAGGCACAATCTTGATAGAAAAGGAGATAATTTTAC 121

122 ATG GGT TGT ACG CTT ATA CTA ATG TTG AAT GTG TTG TTG GTG TTG TTG GGT TCA TGG GTT TTT TCT GGA 190
1 Met Gly Cys Thr Leu Ile Leu Met Leu Asn Val Leu Leu Val Leu Leu Gly Ser Trp Val Phe Ser Gly 23

191 ACA GCT TCT GTT TCA TAT GAC CAT AGG GCT ATT ATT GTA AAT GGA CAA AGA AGA ATA CTT ATT TCT GGT 259
24 Thr Ala Ser Val Ser Tyr Asp His Arg Ala Ile Ile Val Asn Gly Gln Arg Arg Ile Leu Ile Ser Gly 46

260 TCT GTT CAT TAT CCA AGA AGC ACT CCT GAG ATG TGG CCA GGT ATT ATT CAA AAG GCT AAA GAA GGA GGT 328
47 Ser Val His Tyr Pro Arg Ser Thr Pro Glu Met Trp Pro Gly Ile Ile Gln Lys Ala Lys Glu Gly Gly 69

329 GTG GAT GTG ATT CAG ACT TAT GTT TTC TGG AAT GGA CAT GAG CCT CAA CAA GGG AAA TAT TAT TTT GAA 397
70 Val Asp Val Ile Gln Thr Tyr Val Phe Trp Asn Gly His Glu Pro Gln Gln Gly Lys Tyr Tyr Phe Glu 92

398 GGG AGA TAT GAT TTA GTG AAG TTT ATT AAG CTG GTG CAC CAA GCA GGA CTT TAT GTC CAT CTT AGA GTT 466
93 Gly Arg Tyr Asp Leu Val Lys Phe Ile Lys Leu Val His Gln Ala Gly Leu Tyr Val His Leu Arg Val 115

467 GGA CCT TAT GCT TGT GCT GAA TGG AAT TTT GGG GGC TTT CCT GTT TGG CTG AAA TAT GTT CCA GGT ATC 535
116 Gly Pro Tyr Ala Cys Ala Glu Trp Asn Phe Gly Gly Phe Pro Val Trp Leu Lys Tyr Val Pro Gly Ile 138

536 AGT TTC AGA ACA GAT AAT GGA CCT TTC AAG GCT GCA ATG CAA AAA TTT ACT GCC AAG ATT GTC AAT ATG 604
139 Ser Phe Arg Thr Asp Asn Gly Pro Phe Lys Ala Ala Met Gln Lys Phe Thr Ala Lys Ile Val Asn Met 161

605 ATG AAA GCG GAA CGT TTG TAT GAA ACT CAA GGG GGG CCA ATA ATT TTA TCT CAG ATT GAG AAT GAA TAT 673
162 Met Lys Ala Glu Arg Leu Tyr Glu Thr Gln Gly Gly Pro Ile Ile Leu Ser Gln Ile Glu Asn Glu Tyr 184

674 GGA CCC ATG GAA TGG GAA CTG GGA GCA CCA GGT AAA TCT TAC GCA CAG TGG GCC GCC AAA ATG GCT GTG 742
185 Gly Pro Met Glu Trp Glu Leu Gly Ala Pro Gly Lys Ser Tyr Ala Gln Trp Ala Ala Lys Met Ala Val 207

743 GGT CTT GAC ACT GGT GTC CCA TGG GTT ATG TGC AAG CAA GAC GAT GCC CCT GAT CCT ATT ATA AAT GCT 811
208 Gly Leu Asp Thr Gly Val Pro Trp Val Met Cys Lys Gln Asp Asp Ala Pro Asp Pro Ile Ile Asn Ala 230

812 TGC AAT GGC TTC TAC TGT GAC TAC TTT TCT CCA AAC AAG GCT TAT AAA CCA AAG ATA TGG ACT GAA GCC 880
231 Cys Asn Gly Phe Tyr Cys Asp Tyr Phe Ser Pro Asn Lys Ala Tyr Lys Pro Lys Ile Trp Thr Glu Ala 253

881 TGG ACT GCA TGG TTT ACT GGT TTT GGA AAT CCA GTT CCT TAC CGT CCT GCT GAG GAC TTG GCA TTT TCT 949
254 Trp Thr Ala Trp Phe Thr Gly Phe Gly Asn Pro Val Pro Tyr Arg Pro Ala Glu Asp Leu Ala Phe Ser 276

950 GTT GCA AAA TTT ATA CAG AAG GGA GGT TCC TTC ATC AAT TAT TAC ATG TAT CAT GGA GGA ACA AAC TTT 1018
277 Val Ala Lys Phe Ile Gln Lys Gly Gly Ser Phe Ile Asn Tyr Tyr Met Tyr His Gly Gly Thr Asn Phe 299

1019 GGA CGG ACT GCT GGT GGT CCA TTT ATT GCT ACT AGT TAT GAC TAT GAT GCA CCA CTT GAT GAA TAT GGA 1087
300 Gly Arg Thr Ala Gly Gly Pro Phe Ile Ala Thr Ser Tyr Asp Tyr Asp Ala Pro Leu Asp Glu Tyr Gly 322

1088 TTA TTG CGA CAA CCA AAA TGG GGT CAC CTG AAA GAT CTG CAT AGA GCA ATA AAG CTT TGT GAA CCA GCT 1156
323 Leu Leu Arg Gln Pro Lys Trp Gly His Leu Lys Asp Leu His Arg Ala Ile Lys Leu Cys Glu Pro Ala 345

1157 TTA GTC TCT GGA GAT CCA GCT GTG ACA GCA CTT GGA CAC CAG CAG GAG GCC CAT GTT TTT AGG TCG AAG 1225
346 Leu Val Ser Gly Asp Pro Ala Val Thr Ala Leu Gly His Gln Gln Glu Ala His Val Phe Arg Ser Lys 368

1226 GCT GGC TCT TGT GCT GCA TTC CTT GCT AAC TAC GAC CAA CAC TCT TTT GCT ACT GTG TCA TTT GCA AAC 1294
369 Ala Gly Ser Cys Ala Ala Phe Leu Ala Asn Tyr Asp Gln His Ser Phe Ala Thr Val Ser Phe Ala Asn 391

1295 AGG CAT TAC AAC TTG CCA CCA TGG TCA ATC AGC ATT CTT CCC GAC TGC AAG AAC ACT GTA TTT AAT ACA 1363
392 Arg His Tyr Asn Leu Pro Pro Trp Ser Ile Ser Ile Leu Pro Asp Cys Lys Asn Thr Val Phe Asn Thr 414

1364 GCA CGG ATC GGT GCT CAA AGT GCT CAG ATG AAG ATG ACT CCA GTC AGC AGA GGA TTG CCC TGG CAG TCA 1432
415 Ala Arg Ile Gly Ala Gln Ser Ala Gln Met Lys Met Thr Pro Val Ser Arg Gly Leu Pro Trp Gln Ser 437

1433 TTC AAT GAA GAG ACA TCA TCT TAT GAA GAC AGT AGT TTT ACA GTT GTT GGG CTA TTG GAA CAG ATA AAT 1501
438 Phe Asn Glu Glu Thr Ser Ser Tyr Glu Asp Ser Ser Phe Thr Val Val Gly Leu Leu Glu Gln Ile Asn 460

Figure 2
Sheet 6 of 12

Gene/clone name: TBG3/p2-1-3[redacted]bl; accession number AF154421; Seq[redacted]ce ID number 3 cont.

1502 ACA ACA AGA GAC GTG TCT GAT TAT TTG TGG TAT TCA ACA GAT GTC AAG ATT GAT TCA AGA GAA AAG TTT 1570
461 Thr Thr Arg Asp Val Ser Asp Tyr Leu Trp Tyr Ser Thr Asp Val Lys Ile Asp Ser Arg Glu Lys Phe 483

1571 TTG AGA GGC GGA AAA TGG CCT TGG CTT ACG ATC ATG TCA GCT GGG CAT GCA TTG CAT GTT TTT GTG AAT 1639
484 Leu Arg Gly Gly Lys Trp Pro Trp Leu Thr Ile Met Ser Ala Gly His Ala Leu His Val Phe Val Asn 506

1640 GGT CAA TTA GCA GGA ACT GCA TAT GGA AGT TTA GAA AAA CCG AAA CTA ACT TTC AGT AAA GCC GTA AAT 1708
507 Gly Gln Leu Ala Gly Thr Ala Tyr Gly Ser Leu Glu Lys Pro Lys Leu Thr Phe Ser Lys Ala Val Asn 529

1709 CTG AGA GCA GGT GTT AAC AAG ATT TCT CTA CTG AGC ATT GCT GTT GGC CTT CCG AAT ATC GGC CCA CAT 1777
530 Leu Arg Ala Gly Val Asn Lys Ile Ser Leu Leu Ser Ile Ala Val Gly Leu Pro Asn Ile Gly Pro His 552

1778 TTT GAG ACA TGG AAT GCT GGT GTT CTT GGG CCA GTC TCA CTA ACT GGT CTT GAC GAG GGG AAA AGA GAT 1846
553 Phe Glu Thr Trp Asn Ala Gly Val Leu Gly Pro Val Ser Leu Thr Gly Leu Asp Glu Gly Lys Arg Asp 575

1847 TTA ACA TGG CAG AAA TGG TCT TAC AAG GTT GGT CTA AAA GGA GAA GCC TTG AGC CTC CAT TCA CTC AGT 1915
576 Leu Thr Trp Gln Lys Trp Ser Tyr Lys Val Gly Leu Lys Gly Glu Ala Leu Ser Leu His Ser Leu Ser 598

1916 GGT AGC TCG TCA GTT GAG TGG GTC GAG GGT TCT TTA GTG GCT CAG AGA CAG CCA CTC ACA TGG TAC AAG 1984
599 Gly Ser Ser Ser Val Glu Trp Val Glu Gly Ser Leu Val Ala Gln Arg Gln Pro Leu Thr Trp Tyr Lys 621

1985 AGC ACT TTT AAT GCT CCA GCT GGA AAT GAT CCT TTG GCT TTA GAC TTG AAT ACC ATG GGC AAA GGA CAA 2053
622 Ser Thr Phe Asn Ala Pro Ala Gly Asn Asp Pro Leu Ala Leu Asp Leu Asn Thr Met Gly Lys Gly Gln 644

2054 GTG TGG ATA AAT GGT CAA AGC CTC GGA CGC TAT TGG CCT GGA TAT AAA GCA TCT GGT AAC TGC GGT GCC 2122
645 Val Trp Ile Asn Gly Gln Ser Leu Gly Arg Tyr Trp Pro Gly Tyr Lys Ala Ser Gly Asn Cys Gly Ala 667

2123 TGT AAC TAT GCA GGC TGG TTT AAT GAG AAA AAA TGC CTA AGT AAC TGT GGA GAG GCT TCA CAA CGA TGG 2191
668 Cys Asn Tyr Ala Gly Trp Phe Asn Glu Lys Lys Cys Leu Ser Asn Cys Gly Glu Ala Ser Gln Arg Trp 690

2192 TAT CAT GTT CCC CGT TCT TGG CTG TAT CCT ACT GGA AAT TTG TTA GTT CTA TTT GAG GAA TGG GGA GGA 2260
691 Tyr His Val Pro Arg Ser Trp Leu Tyr Pro Thr Gly Asn Leu Leu Val Leu Phe Glu Glu Trp Gly Gly 713

2261 GAG CCT CAT GGA ATC TCT TTG GTA AAA AGA GAA GTT GCA AGT GTT TGT GCA GAT ATA AAC GAA TGG CAA 2329
714 Glu Pro His Gly Ile Ser Leu Val Lys Arg Glu Val Ala Ser Val Cys Ala Asp Ile Asn Glu Trp Gln 736

2330 CCA CAG TTG GTG AAT TGG CAA ATG CAA GCA TCT GGT AAA GTT GAC AAA CCA CTG AGA CCT AAA GCT CAC 2398
737 Pro Gln Leu Val Asn Trp Gln Met Gln Ala Ser Gly Lys Val Asp Lys Pro Leu Arg Pro Lys Ala His 759

2399 CTC TCG TGT GCT TCT GGT CAG AAG ATT ACT TCA ATC AAA TTT GCA AGC TTT GGA ACA CCA CAA GGG GTC 2467
760 Leu Ser Cys Ala Ser Gly Gln Lys Ile Thr Ser Ile Lys Phe Ala Ser Phe Gly Thr Pro Gln Gly Val 782

2468 TGC GGA AGC TTC CGT GAA GGA AGC TGC CAC GCC TTC CAC TCA TAT GAT GCT TTT GAA AGG TAT TGC ATC 2536
783 Cys Gly Ser Phe Arg Glu Gly Ser Cys His Ala Phe His Ser Tyr Asp Ala Phe Glu Arg Tyr Cys Ile 805

2537 GGG CAA AAC TCG TGC TCA GTA CCT GTA ACA CCA GAG ATC TTT GGA GGT GAT CCA TGT CCA CAT GTT ATG 2605
806 Gly Gln Asn Ser Cys Ser Val Pro Val Thr Pro Glu Ile Phe Gly Gly Asp Pro Cys Pro His Val Met 828

2606 AAG AAA CTC TCA GTT GAG GTT ATT TGC AGT TGA TGACACTGAGGAGAAACAAATAAAAGTGGTTTCAGTTAGTTGTCTGAA 2686
829 Lys Lys Leu Ser Val Glu Val Ile Cys Ser *** 840

2687 CATATCAAAAAGTTGGCTTTGATGGAGGTGAAGTTGTACAGATATGCAACACACCTTTCCATTTGAGGCACATATGAATTGCAATGGCCCAA 2778
2779 GATTCTGTACATATATGTTGGTTACTGTCAAGTTGGTATTGGTTTGCAAATGTAAAACAGTAGTATAGTCATTGGTTCAAGTGCGCATCGAG 2870
2871 ATTGTGCTAGTGGGAGGTAGTAGGTACCGATCGATCTATCGTTGTTTGCACAAGCTGGGCCTAGATTCCACTATTATTATAACAAAGAAAGC 2962
2963 ACAATGAGACTGATTCTTGATTAGTCCATGTGTAGATATTGTTACTGTTGGAATTTGCAAATCTTGTGATTTCAGCAAAAAAAAAAAAAAAA 3054
3055 AAAAAAAAAAAAAAA 3069

Figure 2
Sheet 7 of 12

ene/clone name: TBG4/pZBG2-[redacted]pTomβgal4; accession number AF02039[redacted] sequence ID number 4

1 AAAAAAAGTTTCAATTTTTTTTTCTAAAATAAAAAAAAATTCATTTTTTTTGAATGTGGAAAAA 63

64 ATG CTA AGG ACT AAT GTG TTG TTG TTA TTA GTT ATT TGT TTA TTG GAT TTT TTT TCT TCA GTG AAA GCT 132
1 Met Leu Arg Thr Asn Val Leu Leu Leu Leu Val Ile Cys Leu Leu Asp Phe Phe Ser Ser Val Lys Ala 23

133 AGT GTT TCT TAT GAT GAC AGA GCT ATA ATC ATA AAT GGG AAA AGA AAA ATT CTT ATT TCT GGT TCA ATT 201
24 Ser Val Ser Tyr Asp Asp Arg Ala Ile Ile Ile Asn Gly Lys Arg Lys Ile Leu Ile Ser Gly Ser Ile 46

202 CAT TAT CCA AGA AGC ACT CCA CAG ATG TGG CCT GAT CTT ATA CAA AAG GCT AAA GAT GGA GGC TTA GAT 270
47 His Tyr Pro Arg Ser Thr Pro Gln Met Trp Pro Asp Leu Ile Gln Lys Ala Lys Asp Gly Gly Leu Asp 69

271 GTT ATT GAA ACT TAT GTT TTC TGG AAT GGA CAT GAG CCT TCT CCT GGA AAA TAT AAT TTT GAA GGA AGA 339
70 Val Ile Glu Thr Tyr Val Phe Trp Asn Gly His Glu Pro Ser Pro Gly Lys Tyr Asn Phe Glu Gly Arg 92

340 TAT GAT CTT GTT AGA TTC ATC AAA ATG GTA CAA AGA GCA GGA CTT TAT GTC AAT TTA CGT ATT GGC CCT 408
93 Tyr Asp Leu Val Arg Phe Ile Lys Met Val Gln Arg Ala Gly Leu Tyr Val Asn Leu Arg Ile Gly Pro 115

409 TAC GTC TGT GCT GAA TGG AAC TTT GGG GGA TTC CCT GTT TGG CTA AAA TAT GTG CCT GGT ATG GAA TTT 477
116 Tyr Val Cys Ala Glu Trp Asn Phe Gly Gly Phe Pro Val Trp Leu Lys Tyr Val Pro Gly Met Glu Phe 138

478 AGA ACA AAC AAT CAG CCT TTT AAG GTG GCT ATG CAA GGA TTT GTT CAG AAA ATA GTC AAC ATG ATG AAG 546
139 Arg Thr Asn Asn Gln Pro Phe Lys Val Ala Met Gln Gly Phe Val Gln Lys Ile Val Asn Met Met Lys 161

547 TCA GAA AAT TTG TTT GAA TCT CAA GGA GGA CCA ATA ATT ATG GCC CAG ATA GAA AAT GAG TAT GGA CCA 615
162 Ser Glu Asn Leu Phe Glu Ser Gln Gly Gly Pro Ile Ile Met Ala Gln Ile Glu Asn Glu Tyr Gly Pro 184

616 GTA GAA TGG GAA ATT GGT GCT CCT GGT AAA GCT TAT ACA AAA TGG GCA GCT CAA ATG GCT GTA GGT TTG 684
185 Val Glu Trp Glu Ile Gly Ala Pro Gly Lys Ala Tyr Thr Lys Trp Ala Ala Gln Met Ala Val Gly Leu 207

685 AAA ACT GGT GTC CCA TGG ATC ATG TGT AAG CAA GAG GAT GCT CCT GAT CCT GTG ATT GAT ACT TGT AAT 753
208 Lys Thr Gly Val Pro Trp Ile Met Cys Lys Gln Glu Asp Ala Pro Asp Pro Val Ile Asp Thr Cys Asn 230

754 GGC TTC TAC TGC GAA GGG TTC CGT CCT AAT AAG CCT TAC AAA CCT AAA ATG TGG ACA GAA GTA TGG ACT 822
231 Gly Phe Tyr Cys Glu Gly Phe Arg Pro Asn Lys Pro Tyr Lys Pro Lys Met Trp Thr Glu Val Trp Thr 253

823 GGC TGG TAT ACG AAA TTC GGT GGT CCA ATT CCT CAA AGA CCA GCC GAA GAC ATT GCA TTT TCA GTT GCC 891
254 Gly Trp Tyr Thr Lys Phe Gly Gly Pro Ile Pro Gln Arg Pro Ala Glu Asp Ile Ala Phe Ser Val Ala 276

892 AGG TTT GTT CAG AAC AAT GGT TCA TTC TTC AAT TAC TAC ATG TAT CAT GGA GGA ACA AAT TTT GGC CGG 960
277 Arg Phe Val Gln Asn Asn Gly Ser Phe Phe Asn Tyr Tyr Met Tyr His Gly Gly Thr Asn Phe Gly Arg 299

961 ACA TCA TCA GGG CTT TTC ATT GCA ACT AGC TAC GAT TAT GAT GCT CCT CTC GAT GAA TAT GGG TTG CTG 1029
300 Thr Ser Ser Gly Leu Phe Ile Ala Thr Ser Tyr Asp Tyr Asp Ala Pro Leu Asp Glu Tyr Gly Leu Leu 322

1030 AAT GAA CCA AAG TAT GGG CAC TTG AGA GAC TTA CAT AAA GCT ATC AAG CTA TCT GAA CCG GCT TTA GTT 1098
323 Asn Glu Pro Lys Tyr Gly His Leu Arg Asp Leu His Lys Ala Ile Lys Leu Ser Glu Pro Ala Leu Val 345

1099 TCA TCA TAT GCT GCG GTG ACT AGT CTT GGA AGT AAT CAA GAG GCT CAT GTT TAT AGA TCA AAA TCT GGA 1167
346 Ser Ser Tyr Ala Ala Val Thr Ser Leu Gly Ser Asn Gln Glu Ala His Val Tyr Arg Ser Lys Ser Gly 368

1168 GCT TGT GCT GCT TTT TTA TCC AAC TAT GAC TCT AGA TAT TCA GTA AAA GTC ACC TTT CAG AAT AGG CCA 1236
369 Ala Cys Ala Ala Phe Leu Ser Asn Tyr Asp Ser Arg Tyr Ser Val Lys Val Thr Phe Gln Asn Arg Pro 391

1237 TAC AAT CTG CCT CCA TGG TCC ATC AGC ATT CTT CCC GAC TGC AAA ACT GCC GTT TAC AAC ACT GCA CAG 1305
392 Tyr Asn Leu Pro Pro Trp Ser Ile Ser Ile Leu Pro Asp Cys Lys Thr Ala Val Tyr Asn Thr Ala Gln 414

1306 GTT AAC TCT CAA AGC TCG AGC ATA AAG ATG ACG CCT GCA GGT GGT GGA TTG TCT TGG CAG TCA TAC AAT 1374
415 Val Asn Ser Gln Ser Ser Ser Ile Lys Met Thr Pro Ala Gly Gly Gly Leu Ser Trp Gln Ser Tyr Asn 437

1375 GAA GAA ACG CCT ACT GCT GAT GAC AGC GAT ACA CTT ACA GCT AAC GGA CTA TGG GAA CAG AAA AAC GTC 1443
438 Glu Glu Thr Pro Thr Ala Asp Asp Ser Asp Thr Leu Thr Ala Asn Gly Leu Trp Glu Gln Lys Asn Val 460

Figure 2
Sheet 8 of 12

Gene/clone name: TBG4/pZBG2-[redacted]/pTomβgal4; accession number AF0203[redacted] Sequence ID number 4
cont.

1444 ACA AGA GAT TCA TCA GAC TAT CTG TGG TAC ATG ACA AAT GTA AAT ATA GCA TCT AAT GAA GGA TTT CTA 1512
461 Thr Arg Asp Ser Ser Asp Tyr Leu Trp Tyr Met Thr Asn Val Asn Ile Ala Ser Asn Glu Gly Phe Leu 483

1513 AAG AAC GGA AAG GAT CCT TAT CTC ACT GTT ATG TCC GCT GGT CAT GTC TTG CAT GTT TTC GTC AAT GGA 1581
484 Lys Asn Gly Lys Asp Pro Tyr Leu Thr Val Met Ser Ala Gly His Val Leu His Val Phe Val Asn Gly 506

1582 AAA CTA TCA GGA ACT GTT TAT GGT ACA TTG GAT AAT CCA AAA CTT ACA TAC AGT GGC AAC GTG AAG TTA 1650
507 Lys Leu Ser Gly Thr Val Tyr Gly Thr Leu Asp Asn Pro Lys Leu Thr Tyr Ser Gly Asn Val Lys Leu 529

1651 AGA GCT GGT ATT AAC AAG ATT TCT CTG CTC AGT GTT TCC GTT GGT CTC CCG AAC GTT GGC GTG CAT TAT 1719
530 Arg Ala Gly Ile Asn Lys Ile Ser Leu Leu Ser Val Ser Val Gly Leu Pro Asn Val Gly Val His Tyr 552

1720 GAT ACA TGG AAT GCA GGA GTT CTA GGT CCA GTC ACG TTG AGC GGT CTC AAT GAA GGG TCA AGA AAC TTG 1788
553 Asp Thr Trp Asn Ala Gly Val Leu Gly Pro Val Thr Leu Ser Gly Leu Asn Glu Gly Ser Arg Asn Leu 575

1789 GCG AAA CAG AAA TGG TCT TAC AAG GTT GGT CTG AAA GGC GAA TCG TTA AGT CTT CAC TCC TTA AGT GGG 1857
576 Ala Lys Gln Lys Trp Ser Tyr Lys Val Gly Leu Lys Gly Glu Ser Leu Ser Leu His Ser Leu Ser Gly 598

1858 AGT TCT TCT GTT GAA TGG GTT CGA GGT TCA CTA ATG GCT CAA AAG CAG CCC CTG ACT TGG TAC AAG GCT 1926
599 Ser Ser Ser Val Glu Trp Val Arg Gly Ser Leu Met Ala Gln Lys Gln Pro Leu Thr Trp Tyr Lys Ala 621

1927 ACA TTT AAC GCG CCT GGA GGA AAT GAT CCA CTA GCT TTA GAC ATG GCA AGT ATG GGA AAA GGT CAG ATA 1995
622 Thr Phe Asn Ala Pro Gly Gly Asn Asp Pro Leu Ala Leu Asp Met Ala Ser Met Gly Lys Gly Gln Ile 644

1996 TGG ATA AAT GGT GAA GGC GTA GGT CGC CAT TGG CCT GGA TAC ATA GCA CAA GGC GAC TGC AGC AAA TGC 2064
645 Trp Ile Asn Gly Glu Gly Val Gly Arg His Trp Pro Gly Tyr Ile Ala Gln Gly Asp Cys Ser Lys Cys 667

2065 AGT TAT GCT GGA ACG TTC AAC GAG AAG AAG TGC CAG ACT AAC TGC GGA CAA CCT TCT CAG AGA TGG TAC 2133
668 Ser Tyr Ala Gly Thr Phe Asn Glu Lys Lys Cys Gln Thr Asn Cys Gly Gln Pro Ser Gln Arg Trp Tyr 690

2134 CAT GTT CCA CGA TCG TGG CTG AAA CCA AGT GGA AAC TTG TTA GTA GTA TTC GAA GAA TGG GGA GGT AAT 2202
691 His Val Pro Arg Ser Trp Leu Lys Pro Ser Gly Asn Leu Leu Val Val Phe Glu Glu Trp Gly Gly Asn 713

2203 CCA ACA GGA ATT TCT CTA GTC AGG AGA TCA AGA TAA AGAACTCGAAAAGTAAAACTTGTTCAGTAACTATGGTGCTTGAA 2282
714 Pro Thr Gly Ile Ser Leu Val Arg Arg Ser Arg *** 725

2283 TTCGCGCCGAAAAATACATACACGAAGCTAACAATGGAGGCTACAGTTTGCAAATTGCAGCTGAATAAAACATTAGAAGATAAAGAAATATT 2374
2375 TGATTAAAAGGAGTATATAAATTTACAGAGAATTTTCTTTATTCTTTGTAAAACTTTGGTTTATAAAGTTTATACAGAATTTTCTGTTATTT 2466
2467 GGATTATGAGATTGAAGAAGATTGTACAGCTTCCAAATACTATTAGAATACAAATAAATTTCATGTAAAAAAAAAAAAAAAAAAAAAA 2554

Figure 2
She t 9 of 12

Gene/clone name: TBG5/RT[redacted]R2-1/bl; accession number AF154423;[redacted]quence ID number 5

1 ATC CAG ACT TAC GTT TTC TGG AAC CTT CAT GAA CCT GTT CGA AAT CAG TAT GAT TTT GAA GGA AGG AAA 69
1 Ile Gln Thr Tyr Val Phe Trp Asn Leu His Glu Pro Val Arg Asn Gln Tyr Asp Phe Glu Gly Arg Lys 23

70 GAT TTG ATT AAT TTT GTG AAG TTG GTG GAG AGA GCT GGC TTA TTT GTT CAT ATA AGG ATT GGG CCT TAT 138
24 Asp Leu Ile Asn Phe Val Lys Leu Val Glu Arg Ala Gly Leu Phe Val His Ile Arg Ile Gly Pro Tyr 46

139 GTT TGT GCA GAA TGG AAC TAT GGT GGG TTT CCT CTT TGG TTG CAT TTC ATT CCT GGA ATT GAA TTT CGA 207
47 Val Cys Ala Glu Trp Asn Tyr Gly Gly Phe Pro Leu Trp Leu His Phe Ile Pro Gly Ile Glu Phe Arg 69

208 ACC GAC AAT GAA CCG TTC AAG GCA GAA ATG AAG CGA TTC ACA GCT AAA ATT GTT GAC ATG ATC AAG CAA 276
70 Thr Asp Asn Glu Pro Phe Lys Ala Glu Met Lys Arg Phe Thr Ala Lys Ile Val Asp Met Ile Lys Gln 92

277 GAA AAT CTA TAT GCA TCC CAG GGT GGG CCG GTT ATC TTG TCT CAG ATA GAA AAT GAG TAT GGC AAT GGT 345
93 Glu Asn Leu Tyr Ala Ser Gln Gly Gly Pro Val Ile Leu Ser Gln Ile Glu Asn Glu Tyr Gly Asn Gly 115

346 GAT ATT GAG TCT CGT TAT GGT CCT CGT GCC AAA CCT TAC GTG AAC TGG GCA GCA TCA ATG GCT ACG TCT 414
116 Asp Ile Glu Ser Arg Tyr Gly Pro Arg Ala Lys Pro Tyr Val Asn Trp Ala Ala Ser Met Ala Thr Ser 138

415 TTA AAT ACG GGA GTG CCA TGG GTT ATG TGT CAG CAA CCA GAT GCC CCT CCT TCC GTT ATT AAC ACT TGC 483
139 Leu Asn Thr Gly Val Pro Trp Val Met Cys Gln Gln Pro Asp Ala Pro Pro Ser Val Ile Asn Thr Cys 161

484 AAT GGA TTT TAT TGT GAC CAA TTC AAG CAA AAT TCC GAT AAA ACA CCC AAG ATG TGG ACT GAG AAT TGG 552
162 Asn Gly Phe Tyr Cys Asp Gln Phe Lys Gln Asn Ser Asp Lys Thr Pro Lys Met Trp Thr Glu Asn Trp 184

553 ACC GGA TGG TTT CTG TCG TTT GGT GGT CCT GTC CCT TAC AGA CCA GTG GAA GAC ATC GCT TTC GCT GTG 621
185 Thr Gly Trp Phe Leu Ser Phe Gly Gly Pro Val Pro Tyr Arg Pro Val Glu Asp Ile Ala Phe Ala Val 207

622 GCT CGA TTT TTC CAG CGA GGC GGA ACT TTC CAG AAC TAT TAC ATG TAC CAC GGG GGA ACT AAC TTT GGG 690
208 Ala Arg Phe Phe Gln Arg Gly Gly Thr Phe Gln Asn Tyr Tyr Met Tyr His Gly Gly Thr Asn Phe Gly 230

691 AGA ACC AGT GGT GGA CCG TTT ATT GCA ACT AGC TAT GAC TAT GAT GCC CCT CTC GAC GAA TAC GG 755
231 Arg Thr Ser Gly Gly Pro Phe Ile Ala Thr Ser Tyr Asp Tyr Asp Ala Pro Leu Asp Glu Tyr 252

Figure 2
Sheet 10 of 12

Gene/clone name: TBG6/RT[redacted]R2-6/bl; accession number AF154424;[redacted]quence ID number 6

```
1   ATC CAG ACA TAT GTT TTT TGG AAT GTT CAT GAG CCT TCT CCT GGC AAT TAC AAT TTT GAA GGA AGA TAT   69
1   Ile Gln Thr Tyr Val Phe Trp Asn Val His Glu Pro Ser Pro Gly Asn Tyr Asn Phe Glu Gly Arg Tyr   23

70  GAC CTG GTG AGG TTT GTA AAA ACG ATT CAG AAA GCA GGG CTG TAT GCT CAT CTT CGA ATT GGC CCT TAC   138
24  Asp Leu Val Arg Phe Val Lys Thr Ile Gln Lys Ala Gly Leu Tyr Ala His Leu Arg Ile Gly Pro Tyr   46

139 GTT TGT GCA GAG TGG AAT TTT GGA GGG TTT CCA GTA TGG CTG AAG TAT GTA CCT GGC ATT AGC TTC AGA   207
47  Val Cys Ala Glu Trp Asn Phe Gly Gly Phe Pro Val Trp Leu Lys Tyr Val Pro Gly Ile Ser Phe Arg   69

208 GCT GAT AAT GAA CCT TTC AAG AAC GCA ATG AAA GGG TAT GCT GAG AAA ATT GTT AAC TTG ATG AAG ATC   276
70  Ala Asp Asn Glu Pro Phe Lys Asn Ala Met Lys Gly Tyr Ala Glu Lys Ile Val Asn Leu Met Lys Ile   92

277 ATA ATC TTT TCG AGT CTC AGG GTG GTC CAA TCA TAC TCT CAC AGA TTG AGA ATG AGT ATG GGC CTC AAG   345
93  Ile Ile Phe Ser Ser Leu Arg Val Val Gln Ser Tyr Ser His Arg Leu Arg Met Ser Met Gly Leu Lys   115

346 CCA AGG TAC TTG GAG CAC CGG GAC ATC AGT ATT CAA CAT GGG CTG CAA ATA TGG CAG TTG GAT TTG AAC   414
116 Pro Arg Tyr Leu Glu His Arg Asp Ile Ser Ile Gln His Gly Leu Gln Ile Trp Gln Leu Asp Leu Asn   138

415 ACA GGC GTC CCA TGG GTG ATG TGC AAG GAA GAA GAT GCA CCA GAT CCT GTG ATC AAC ACA TGC AAT GGT   483
139 Thr Gly Val Pro Trp Val Met Cys Lys Glu Glu Asp Ala Pro Asp Pro Val Ile Asn Thr Cys Asn Gly   161

484 TTC TAC TGT GAT AAT TTC TTC CCA AAC AAA CCA TAC AAA CCT GCA ATT TGG ACT GAA GCT TGG AGT GGA   552
162 Phe Tyr Cys Asp Asn Phe Phe Pro Asn Lys Pro Tyr Lys Pro Ala Ile Trp Thr Glu Ala Trp Ser Gly   184

553 TGG TTC TCG GAA TTT GGC GGT CCC CTT CAT CAG AGA CCA GTT CAG GAT TTG GCA TTT GCT GTT GCC CAA   621
185 Trp Phe Ser Glu Phe Gly Gly Pro Leu His Gln Arg Pro Val Gln Asp Leu Ala Phe Ala Val Ala Gln   207

622 TTT ATA CAA AGA GGA GGA TCT TTT GTT AAC TAT TAC ATG TAC CAT GGG GGC ACG AAC TTT GGA CGC ACT   690
208 Phe Ile Gln Arg Gly Gly Ser Phe Val Asn Tyr Tyr Met Tyr His Gly Gly Thr Asn Phe Gly Arg Thr   230

691 GCG GGT GGG CCA TTC ATC ACT ACC AGC TAT GAT TAT GAT GCC CCC CTC GAC GAG TAT GG                749
231 Ala Gly Gly Pro Phe Ile Thr Thr Ser Tyr Asp Tyr Asp Ala Pro Leu Asp Glu Tyr                   250
```

Figure 2

Sheet 11 of 12

Gene/clone name: TBG7/[redacted]G2-1-18; accession number AF154422; [redacted] number 7

1 GCAACTCTCTCC 12

13 GTGAATAACACCGGTAAACGGCCAATGCCAACTCTCGTCGGAATCTGAATAGTGATTTAAGCAGCTTAGCTAGCTAACTTTTGCCTCTGCA 103

104 ATG AAC ACA ATG AGT TGT TTG TCC TCT AAT TTC AAG TTC GTT TTC CTT GCC TCG ACT GTG ATA TGG ATG 172

1 Met Asn Thr Met Ser Cys Leu Ser Ser Asn Phe Lys Phe Val Phe Leu Ala Ser Thr Val Ile Trp Met 23

173 ACG GTA ATG TCG TCG TCG TTA GCA GCA GTA GAT GCT TCC AAT GTT ACT ACT ATT GGT ACT GAT AGT GTG 241

24 Thr Val Met Ser Ser Ser Leu Ala Ala Val Asp Ala Ser Asn Val Thr Thr Ile Gly Thr Asp Ser Val 46

242 ACT TAC GAT CGA CGC TCG TTG ATT ATT AAC GGC CAG AGG AAG CTG CTC ATC TCC GCT TCC ATT CAC TAT 310

47 Thr Tyr Asp Arg Arg Ser Leu Ile Ile Asn Gly Gln Arg Lys Leu Leu Ile Ser Ala Ser Ile His Tyr 69

311 CCT CGC AGT GTC CCT GCC ATG TGG CCT GGT CTG GTT CGA TTG GCG AAG GAA GGA GGA GTG GAT GTT ATT 379

70 Pro Arg Ser Val Pro Ala Met Trp Pro Gly Leu Val Arg Leu Ala Lys Glu Gly Gly Val Asp Val Ile 92

380 GAA ACG TAT GTT TTC TGG AAC GGT CAC GAA CCT TCT CCG GGC AAT TAT TAC TTT GGA GGA AGG TTT GAT 448

93 Glu Thr Tyr Val Phe Trp Asn Gly His Glu Pro Ser Pro Gly Asn Tyr Tyr Phe Gly Gly Arg Phe Asp 115

449 CTA GTC AAA TTT TGT AAG ATC ATT CAG CAG GCT GGA ATG TAT ATG ATT CTT CGG ATT GGA CCA TTT GTA 517

116 Leu Val Lys Phe Cys Lys Ile Ile Gln Gln Ala Gly Met Tyr Met Ile Leu Arg Ile Gly Pro Phe Val 138

518 GCT GCA GAA TGG AAC TTT GGT GGA CTT CCT GTG TGG TTG CAT TAT GTG CCA GGT ACC ACC TTT CGG ACT 586

139 Ala Ala Glu Trp Asn Phe Gly Gly Leu Pro Val Trp Leu His Tyr Val Pro Gly Thr Thr Phe Arg Thr 161

587 GAT AGT GAA CCA TTT AAG TAT CAC ATG CAG AAG TTC ATG ACA TAT ACA GTG AAC TTA ATG AAG AGA GAG 655

162 Asp Ser Glu Pro Phe Lys Tyr His Met Gln Lys Phe Met Thr Tyr Thr Val Asn Leu Met Lys Arg Glu 184

656 AGG CTT TTT GCA TCT CAA GGA GGT CCA ATC ATC TTG TCA CAG GTA GAA AAT GAG TAC GGC TAC TAT GAA 724

185 Arg Leu Phe Ala Ser Gln Gly Gly Pro Ile Ile Leu Ser Gln Val Glu Asn Glu Tyr Gly Tyr Tyr Glu 207

725 AAT GCA TAT GGA GAA GGA GGG AAA AGG TAT GCC TTA TGG GCT GCT AAA ATG GCC CTT TCT CAA AAT ACT 793

208 Asn Ala Tyr Gly Glu Gly Gly Lys Arg Tyr Ala Leu Trp Ala Ala Lys Met Ala Leu Ser Gln Asn Thr 230

794 GGT GTA CCT TGG ATA ATG TGC CAG CAG TAT GAT GCT CCT GAT CCT GTG ATT GAC ACA TGC AAT TCA TTT 862

231 Gly Val Pro Trp Ile Met Cys Gln Gln Tyr Asp Ala Pro Asp Pro Val Ile Asp Thr Cys Asn Ser Phe 253

863 TAC TGC GAC CAA TTT AAA CCA ATC TCT CCA AAC AAG CCC AAA ATT TGG ACA GAG AAC TGG CCG GGA TGG 931

254 Tyr Cys Asp Gln Phe Lys Pro Ile Ser Pro Asn Lys Pro Lys Ile Trp Thr Glu Asn Trp Pro Gly Trp 276

932 TTC AAG ACA TTT GGG GCC AGA GAT CCT CAC AGG CCT GCA GAA GAT GTT GCT TAT TCC GTG GCT CGT TTT 1000

277 Phe Lys Thr Phe Gly Ala Arg Asp Pro His Arg Pro Ala Glu Asp Val Ala Tyr Ser Val Ala Arg Phe 299

1001 TTC CAA AAA GGA GGA AGC GTG CAG AAT TAT TAC ATG TAC CAT GGT GGG ACG AAC TTT GGC AGG ACA GCA 1069

300 Phe Gln Lys Gly Gly Ser Val Gln Asn Tyr Tyr Met Tyr His Gly Gly Thr Asn Phe Gly Arg Thr Ala 322

1070 GGT GGC CCT TTC ATT ACC ACA AGT TAT GAC TAT GAT GCC CCA ATT GAC GAA TAT GGT TTA CCA AGG TTT 1138

323 Gly Gly Pro Phe Ile Thr Thr Ser Tyr Asp Tyr Asp Ala Pro Ile Asp Glu Tyr Gly Leu Pro Arg Phe 345

1139 CCA AAA TGG GGT CAC CTT AAA GAA CTT CAT AAA GTC ATA AAA TCG TGT GAG CAT GCT CTG CTG AAC AAT 1207

346 Pro Lys Trp Gly His Leu Lys Glu Leu His Lys Val Ile Lys Ser Cys Glu His Ala Leu Leu Asn Asn 368

1208 GAT CCA ACT CTT CTT TCA TTA GGT CCT CTA CAA GAG GCT GAT GTT TAT GAA GAT GCT TCA GGC GCT TGT 1276

369 Asp Pro Thr Leu Leu Ser Leu Gly Pro Leu Gln Glu Ala Asp Val Tyr Glu Asp Ala Ser Gly Ala Cys 391

1277 GCT GCC TTT CTC GCG AAT ATG GAT GAC AAA AAT GAC AAG GTG GTA CAG TTC CGA CAT GTA TCA TAC CAC 1345

392 Ala Ala Phe Leu Ala Asn Met Asp Asp Lys Asn Asp Lys Val Val Gln Phe Arg His Val Ser Tyr His 414

1346 TTG CCA GCA TGG TCT GTT AGC ATT TTG CCA GAC TGC AAA AAT GTA GCG TTC AAC ACA GCA AAG GTT GGA 1414

415 Leu Pro Ala Trp Ser Val Ser Ile Leu Pro Asp Cys Lys Asn Val Ala Phe Asn Thr Ala Lys Val Gly 437

1415 TGT CAA ACT TCT ATT GTC AAT ATG GCA CCC ATA GAT TTG CAT CCC ACC GCA AGT TCA CCA AAG AGA GAC 1483

438 Cys Gln Thr Ser Ile Val Asn Met Ala Pro Ile Asp Leu His Pro Thr Ala Ser Ser Pro Lys Arg Asp 460

Figur 2
Sheet 12 of 12
Gene/clone name: TBG7/pZBG[redacted]-18; accession number AF154422; Seq[redacted]ce ID number 7 cont.

1484 ATC AAG TCT CTT CAG TGG GAA GTC TTC AAG GAA ACA GCT GGA GTA TGG GGA GTT GCT GAT TTC ACT AAA 1552
461 Ile Lys Ser Leu Gln Trp Glu Val Phe Lys Glu Thr Ala Gly Val Trp Gly Val Ala Asp Phe Thr Lys 483

1553 AAC GGA TTT GTA GAT CAC ATT AAC ACC ACA AAA GAT GCT ACA GAC TAC CTC TGG TAC ACA ACA AGT ATT 1621
484 Asn Gly Phe Val Asp His Ile Asn Thr Thr Lys Asp Ala Thr Asp Tyr Leu Trp Tyr Thr Thr Ser Ile 506

1622 TTT GTT CAT GCA GAG GAG GAT TTC CTA AGA AAC AGA GGC ACT GCA ATG CTT TTC GTT GAA TCA AAG GGT 1690
507 Phe Val His Ala Glu Glu Asp Phe Leu Arg Asn Arg Gly Thr Ala Met Leu Phe Val Glu Ser Lys Gly 529

1691 CAT GCT ATG CAT GTC TTC ATC AAT AAA AAG CTT CAA GCC AGT GCA TCT GGA AAT GGC ACA GTG CCA CAG 1759
530 His Ala Met His Val Phe Ile Asn Lys Lys Leu Gln Ala Ser Ala Ser Gly Asn Gly Thr Val Pro Gln 552

1760 TTC AAG TTT GGA ACT CCT ATT GCT CTA AAG GCA GGG AAG AAT GAA ATT TCC TTG TTA AGC ATG ACT GTG 1828
553 Phe Lys Phe Gly Thr Pro Ile Ala Leu Lys Ala Gly Lys Asn Glu Ile Ser Leu Leu Ser Met Thr Val 575

1829 GGC CTA CAA ACA GCT GGA GCG TTT TAT GAA TGG ATT GGA GCT GGT CCA ACA AGT GTC AAA GTT GCA GGG 1897
576 Gly Leu Gln Thr Ala Gly Ala Phe Tyr Glu Trp Ile Gly Ala Gly Pro Thr Ser Val Lys Val Ala Gly 598

1898 TTC AAG ACT GGG ACT ATG GAC TTG ACT GCG TCT GCT TGG ACC TAT AAG ATT GGA TTG CAA GGA GAA CAT 1966
599 Phe Lys Thr Gly Thr Met Asp Leu Thr Ala Ser Ala Trp Thr Tyr Lys Ile Gly Leu Gln Gly Glu His 621

1967 TTG AGG ATA CAG AAG TCA TAT AAC TTG AAG AGT AAA ATT TGG GCA CCA ACT TCG CAG CCA CCA AAG CAA 2035
622 Leu Arg Ile Gln Lys Ser Tyr Asn Leu Lys Ser Lys Ile Trp Ala Pro Thr Ser Gln Pro Pro Lys Gln 644

2036 CAG CCC CTC ACA TGG TAT AAG GCA GTA GTA GAT GCG CCT CCT GGT AAT GAA CCT GTT GCA CTT GAT ATG 2104
645 Gln Pro Leu Thr Trp Tyr Lys Ala Val Val Asp Ala Pro Pro Gly Asn Glu Pro Val Ala Leu Asp Met 667

2105 ATT CAT ATG GGA AAA GGA ATG GCT TGG TTG AAT GGA CAA GAA ATT GGC AGA TAT TGG CCG AGG AGA ACT 2173
668 Ile His Met Gly Lys Gly Met Ala Trp Leu Asn Gly Gln Glu Ile Gly Arg Tyr Trp Pro Arg Arg Thr 690

2174 TCT AAA TAT GAG AAT TGT GTT ACT CAA TGT GAC TAC AGA GGC AAA TTT AAC CCT GAT AAG TGT GTC ACT 2242
691 Ser Lys Tyr Glu Asn Cys Val Thr Gln Cys Asp Tyr Arg Gly Lys Phe Asn Pro Asp Lys Cys Val Thr 713

2243 GGC TGT GGA CAA CCT ACA CAG AGA TGG TAT CAT GTG CCA CGA TCT TGG TTC AAG CCA TCA GGA AAT GTC 2311
714 Gly Cys Gly Gln Pro Thr Gln Arg Trp Tyr His Val Pro Arg Ser Trp Phe Lys Pro Ser Gly Asn Val 736

2312 TTA ATT ATC TTT GAG GAA ATA GGT GGA GAT CCC TCT CAA ATT AGA TTC TCA ATG CGA AAG GTT TCT GGA 2380
737 Leu Ile Ile Phe Glu Glu Ile Gly Gly Asp Pro Ser Gln Ile Arg Phe Ser Met Arg Lys Val Ser Gly 759

2381 GCT TGT GGT CAT CTT TCA GTG GAC CAT CCA TCC TTT GAT GTT GAA AAT CTG CAA GGA AGT GAA ATT GAG 2449
760 Ala Cys Gly His Leu Ser Val Asp His Pro Ser Phe Asp Val Glu Asn Leu Gln Gly Ser Glu Ile Glu 782

2450 AAC GAC AAA AAC AGG CCA ACT CTA AGT TTG AAA TGC CCC ACA AAT ACT AAT ATT TCC TCT GTC AAA TTT 2518
783 Asn Asp Lys Asn Arg Pro Thr Leu Ser Leu Lys Cys Pro Thr Asn Thr Asn Ile Ser Ser Val Lys Phe 805

2519 GCC AGC TTT GGA AAT CCT AAT GGT ACA TGT GGC TCC TAC ATG CTA GGA GAC TGC CAC GAT CAG AAT TCT 2587
806 Ala Ser Phe Gly Asn Pro Asn Gly Thr Cys Gly Ser Tyr Met Leu Gly Asp Cys His Asp Gln Asn Ser 828

2588 GCA GCA CTG GTC GAA AAG GTT TGC CTG AAC CAA AAT GAG TGT GCA TTA GAA ATG TCC AGC GCA AAC TTT 2656
829 Ala Ala Leu Val Glu Lys Val Cys Leu Asn Gln Asn Glu Cys Ala Leu Glu Met Ser Ser Ala Asn Phe 851

2657 AAC ATG CAA TTG TGT CCA AGT ACA GTA AAG AAA CTT GCA GTT GAA GTG AAT TGC AGC TGA GTGTCATTGCCC 2728
852 Asn Met Gln Leu Cys Pro Ser Thr Val Lys Lys Leu Ala Val Glu Val Asn Cys Ser *** 871

2729 AAAATGAATGACATATTCTAATTTATATAGTTTGCTACGGAGATGCTCATTCTTAAACCTTTCTTATATAGCAGAAAAATCTGCTATTCCTT 2820
2821 CTTTCGTCTATGATTTGAAGTTTAAGATATGAGTACTGATGTCTTATTAAGCATCACCAGATAACCTTGGATATTCATGTTTGAAAGACTAA 2912
2913 GTATTCATATTTATTCAGTCGAGATGCAAGATTTATTTGTGAAAAAAAAAAAAAAAAAAA 2972

DNASIS
Multiple Edit1

Figure 3
Sheet 1 of 4

```
                           10         20         30         40         50
TBG1-ORF      -24 .......... .......... ....MGFWMA MLLMLLLCLW VSCGISVSYD    26
TBG2-ORF      -14 .......... ....MSRRKT LNFPLLLTVL TIHFVTVAGE YFKPFNVTYD    36
TBG3-ORF      -20 .......... .......... MGCTLLLMLN VLLVLLGSWV FSGTASVSYD    30
TBG4-ORF      -22 .......... .......... ..MLRTNVLL LLVICLLDFF SSVKASVSYD    28
TBG5-ORF        1 ---------- ---------- ---------- ---------- ----------    50
TBG6-ORF        1 ---------- ---------- ---------- ---------- ----------    50
TBG7-ORF       -1 .MNTMSCLSS NFKFVFLAST VIWMTVMSSS LAAVDASNVT TIGTDSVTYD    49
apple         -21 .......... .......... .MGVGIQTMW SILLLFSCIF SAASASVSYD    29
carnation     -16 .......... ......MLCG KENNVMKMML VYVFVLITLI SCVYGNVWYD    34
asparagus     -20 .......... .......... MALKLVLMLM VALLAAVWSP PAVTASVTYD    30
broccoli      -20 .......... .......... MKMKQFNLLS LFLILITSFG SANSTIVSHD    30
Lupin         -12 .......... ..MFGSRIVM ESLMSRRNFH MVLLLLFFWV CYVTASVTYD    38

                           60         70         80         90        100
TBG1-ORF       27 HKALIVNGQR KLLISGSIHY PRSTPEMWPD LIQKAKEGGV DVIQTYVFWN    76
TBG2-ORF       37 NRALLIGGKR RMLISAGIHY PRATPEMWPT LIARSKEGGA DVIETYTFWN    86
TBG3-ORF       31 HRAITVNGQR RILISGSVHY PRSTPEMWPG IIQKAKEGGV DVIQTYVFWN    80
TBG4-ORF       29 DRALLINGKR KILISGSIHY PRSTRQMWPD LIQKAKDGGL DVIETYVFWN    78
TBG5-ORF       51 ---------- ---------- ---------- ---------- --KQTRAFWN   100
TBG6-ORF       51 ---------- ---------- ---------- ---------- --IQTYVFWN   100
TBG7-ORF       50 RRSLIINGQR KLLISASIHY PRSVPAMWPG LVRLAKEGGV DVIETYVFWN    99
apple          30 HKALLINGQK RILISGSIHY PRSTPEMWPD LIQKAKDGGL DVIQTYVFWN    79
carnation      35 YRAIKINDQR RILLSGSIHY PRSTPEMWPD IIEKAKDSQL DVIQTYVFWN    84
asparagus      31 HKSVLINGQR RILISGSIHY PRSTPEMWPD LIQKAKDGGL DVIQTYVFWN    80
broccoli       31 ERAITIDGQR RLLLSGSIHY PRSTSDMWPD LISKAKDGGL DTIETYVFWN    80
Lupin          39 HKAIMINGQR RILISGSIHY PRSTRQMWPD LIQKAKDGGL DVIETYVFWN    88

                          110        120        130        140        150
TBG1-ORF       77 GHEPEEGKYY FEERYDHVKF EKVVQEAGLY VHLRIGPYAC AEWNFGGFPV   126
TBG2-ORF       87 GHEPTRGQYN FEGRYDLIKF AKEVGSHGLF LFIRIGPYAC AEWNFGGFPI   136
TBG3-ORF       81 GHEPQQGKYY FEGRYDLVKF IKLIHEAGLY VHLRVGPYAC AEWNFGGFPV   130
TBG4-ORF       79 GHEPSPGKYN FEGRYDLVRF LKMVQRAGLY VNLRIGPYVC AEWNFGGFPV   128
TBG5-ORF      101 LHEPVRNQYD FEGRKDLINF VKLVERAGLF VHLRIGPYVC AEWNYGGFPL   150
TBG6-ORF      101 VHEPSPGNYN FEGRYDLVRF VKTIQKAGLY AHLRIGPYVC AEWNFGGFPV   150
TBG7-ORF      100 GHEPSPGNYY FGGRFDLVKF CKIIQQAGMY MILRIGPFVA AEWNFGSLPV   149
apple          80 GHEPSPGNYY FEERYDLVKF IKLVQQEGLF VNLRIGPYVC AEWNFGGFPV   129
carnation      85 GHEPSEGKYY FEGRYDLVKF IKLIHQAGLF VHLRIGPFAC AEWNFGGFPV   134
asparagus      81 GHEPSPGQYY FGGRYDLVRF LKLVKQAGLY AHLRIGPYVC AEWNFGGFPV   130
broccoli       81 AHEPSRRQYD FSGNLDLVRF IKTIQSAGLY SVLRIGPYVC AEWNYGGFPV   130
Lupin          89 GHEPSPGKYY FEDRFDLVGF IKLVQQAGLF VHLRIGPFIC AEWNFGGFPV   138

                          160        170        180        190        200
TBG1-ORF      127 WLKYVPGISF RTNNEPFKAA MQKFTTKIVD MMK-----AE KLYETQGGPI   176
TBG2-ORF      137 WLRDIPGIEF RTDNAPFKEE MERYVKKIVD LMI-----SE SLFSWQGGPI   186
TBG3-ORF      131 WLKYVPGISF RTDNGPFKAA MQKFTAKIVN MMK-----AE RLYETQGGPI   180
TBG4-ORF      129 WLKYVPGMEF RTNNQPFKVA MQGFVQKIVN MMK-----SE NLFESQGGPI   178
TBG5-ORF      151 WLHFIPGIEF RTDNEPFKAE MKRFTAKIVD MIK-----QE NLYASQGGPV   200
TBG6-ORF      151 WLKYVPGISF RADNEPFKNA MKGYAEKIVN LMKIIIFSSL RVVQSYSHRL   200
TBG7-ORF      150 WLHYVPGTTF RTDSEPFKYH MQKFMTYTVN LMK-----RE RLFASQGGPI   199
apple         130 WLKYVPGIAF RTDNEPFKAA MQKFTEKIVS MMK-----AE KLFQTQGGPI   179
carnation     135 WLKYVPGIEF RTDNGPFKEK MQVFTTKIVD MMK-----AE KLFHWQGGPI   184
asparagus     131 WLKYVPGIHF RTDNGPFKAA MGKFTEKIVS MMK-----AE GLYETQGGPI   180
broccoli      131 WLHNMPDMKF RTINPGFMNE MQNFTTKIVN MMK-----EE SLFASQGGPI   180
Lupin         139 WLKYVPGIAF RTDNEPFKEA MQKFTEKIVN IMK-----AE KLFQSQGGPI   188

                          210        220        230        240        250
TBG1-ORF      177 ILSQ-IENEY GP--MEWELG EPGKVYSEWA AKMAVDLGTG VPWIMCKQD-   226
TBG2-ORF      187 ILLQ-IENEY GN--VESSFG PKGKLYMKWA AEMAVGLGAG VPWVMCRQ-T   236
TBG3-ORF      181 ILSQ-IENEY GP--MEWELG APGKSYAQWA AKMAVGLDTG VPWVMCKQD-   230
TBG4-ORF      179 IMAQ-IENEY GP--VEWEIG APGKAYTKWA AQMAVGLKTG VPWIMCKQE-   228
TBG5-ORF      201 ILSQ-IENEY GNGDIESRYG PRAKPYVNWA ASMATSLNTG VPWVMCQQ-P   250
TBG6-ORF      201 RMSMGLKPRY ----LEHRDI SIQHGLQIWQ ----LDLNTG VPWVMCKEE-   250
TBG7-ORF      200 ILSQ-VENEY G--YYENAYG EGGKRYALWA AKMALSQNTG VPWIMC-QQY   249
apple         180 ILSQ-IENEF GP--VEWEIG APGKAYTKWA AQMAVGLDTG VPWIMCKQE-   229
carnation     185 ILNQ-IENEY GP--VEWEIG APGKAYTHWA AQMAQSLNAG VPWIMCKQDS   234
```

DNASIS
Multiple Edit1

Figure 3
Sheet 2 f 4

asparagus	181	ELSQ-IENEY	GP--VEYYDG	AAGKSYTNWA	AKMAVGLNTG	VPWVMCKQD-	230
broccoli	181	ELAQ-IENEY	GN--VISSYG	AEGKAYIDWC	ANMANSLDIG	VPWIMC-QQP	230
Lupin	189	ELSQ-IENEY	GP--VEWEIG	APGKAYTKWA	AGMAVGLDTG	VPWVMCKQE-	238

		260	270	280	290	300	
TBG1-ORF	227	DVPDPIINTC	NGFYCDYFTP	NKANKPKMWT	EAWTAWFTEF	GGPVPYRPAE	276
TBG2-ORF	237	DAPEYIIDTC	NAYYCDGFTP	NSEKKPKIWT	ENWNGWFADW	GERLPYRPSE	286
TBG3-ORF	231	DAPDPIINAC	NGFYCDYFSP	NKAYKPKIWT	EAWTAWFTGF	GNPVPYRPAE	280
TBG4-ORF	229	DAPDPVIDTC	NGFYCEGFRP	NKPYKPKMWT	EVWTGWYTKF	GGPIPQRPAE	278
TBG5-ORF	251	DAPPSVINTC	NGFYCDQFKQ	NSDKTPKMWT	ENWTGWFLSF	GGPVPYRPVE	300
TBG6-ORF	251	DAPDPVINTC	NGFYCDNFFP	NKPYKPAIWT	EAWSGWFSEF	GGPLHQRPVQ	300
TBG7-ORF	250	DAPDPVIDTC	NSFYCDQFKP	ISPNKPKIWT	ENWPGWFKTF	GARDPHRPAE	299
apple	230	DAPDPVIDTC	NGFYCENFKP	NKDYKPKMWT	EVWTGWYTEF	GGAVPTRPAE	279
carnation	235	DVPDNVIDTC	NGFYCEGFVP	KDKSKPKMWT	ENWTGWYTEY	GKPVPYRPAE	284
asparagus	231	DAPDPVINTC	NGFYCDYFSP	NKDNKPKMWT	EAWTGWFTGF	GGAVPQRPAE	280
broccoli	231	HAPQPMIETC	NGFYCDQYKP	SNPSSPKMWT	ENWTGWFKNW	GGKHPYRTAE	280
Lupin	239	DAPDPIIDTC	NGFYCENFTP	NKNYKPKLWT	ENWTGWYTAF	GSATPYRPAE	288

		310	320	330	340	350	
TBG1-ORF	277	DMAFAVARFI	QTGGSFINYY	MYHGGTNFGR	TSGGP[illegible]	[illegible]DAPIDEF	326
TBG2-ORF	287	DEAFAIARFF	QRGGSLQNYY	MYFGGTNFGR	TAGGPTQITS	[illegible]YDAPIDEY	336
TBG3-ORF	281	DLAFSVAKFI	QKGGSFINYY	MYHGGTNFGR	TAGGPF[illegible]	[illegible]DEY	330
TBG4-ORF	279	DLAFSVARFV	QNNGSFFNYY	MYHGGTNFGR	TSSGL[illegible]	[illegible]DEY	328
TBG5-ORF	301	DEAFAVARFF	QRGGTFQNYY	MYHGGTNFGR	TSGGPF[illegible]	[illegible]DEY	350
TBG6-ORF	301	DLAFAVAQFI	QRGGSFVNYY	MYHGGTNFGR	TAGGPFITTS	[illegible]DEY	350
TBG7-ORF	300	DVAYSVARFF	QKGGSVQNYY	MYHGGTNFGR	TAGGPFITTS	[illegible]PIDEY	349
apple	280	DVAFSVARFI	QSGGSFLNYY	MYHGGTNFGR	TAGGPFMATS	[illegible]DEY	329
carnation	285	DVAFSVARFI	QNGGSFMNYY	MFHGGTNFE-	TTAGRFVSTS	[illegible]	334
asparagus	281	DMAFAVARFI	QKGGSFINYY	MYHGGTNFGR	TAGGPFISTS	[illegible]DAPIDEY	330
broccoli	281	DLAFSVARFF	QTGGTFQNYY	MYHGGTNFGR	VAGGPYITTS	[illegible]YDAPIDEY	330
Lupin	289	DLAFSVARFI	QNRGSLFNYY	MYHGGTNFGR	TSNGLEVATS	[illegible]YDAPIDEY	338

		360	370	380	390	400	
TBG1-ORF	327	GSLRQPKWGH	LKDLHRAIKL	CEPALVSVD-	PIVTSLGNYQ	EARVFKSES-	376
TBG2-ORF	337	GELRQPKWGH	LKDLHAAIKL	CEPALVAADS	PQYIKLGPKQ	EAHVARGTSN	386
TBG3-ORF	331	GMLRQPKWGH	LKDLHRAIKL	CEPALVSGD-	PAVTALGHQQ	EAHVFKSKA-	380
TBG4-ORF	329	GLLNEPKYGH	LRDLHKAIKL	SEPALVSSY-	AAVTSLGSNQ	EAHVYRSKS-	378
TBG5-ORF	351	----------	----------	----------	----------	----------	400
TBG6-ORF	351	----------	----------	----------	----------	----------	400
TBG7-ORF	350	GLPRFPKWGH	LKELHKVIKS	CEHALLNND-	PTLLSLGPLQ	EADVYEDAS-	399
apple	330	GLPREPKWGH	LRDLHKAIKS	CESALVSVD-	PSVIKLGSNQ	EAHVFKSES-	379
carnation	335	GLPREPKYTH	LKNLHKAIKM	CEPALVSSD-	AKVTNLGSNQ	EAHVYSSNS-	384
asparagus	331	GLLRQPKWGH	LRDLHKAIKL	CEPALVSGE-	PTITSLGQNQ	ESYVYRSKS-	380
broccoli	331	GNLNQPKWGH	LKQLHTLLKS	MEKPLTYGNI	STID-LGNSV	TATVYSTNEK	380
Lupin	339	GLLNEPKWGH	LRELHRAIKQ	CESALVSVD-	PTVSWPGKNL	EVHLYKTES-	388

		410	420	430	440	450	
TBG1-ORF	377	----------	GACAAFLANY	NQHSFAKVAF	GNMHYNLPPW	SISILPDCKN	426
TBG2-ORF	387	NIGQYMSLNE	GICAAFIANI	DEHESATVKF	YGQEFTLPPW	SVVF---CQI	436
TBG3-ORF	381	----------	GSCAAFLANY	DQHSFATVSF	ANRHYNLPPW	SISILPDCKN	430
TBG4-ORF	379	----------	GACAAFLSNY	DSRYSVKVTF	QNRPYNLPPW	SISILPDCKT	428
TBG5-ORF	401	----------	----------	----------	----------	----------	450
TBG6-ORF	401	----------	----------	----------	----------	----------	450
TBG7-ORF	400	----------	GACAAFLANM	DDKNDKVVQF	RHVSYHLPAW	SVSILPDCKN	449
apple	380	----------	D-CAAFLANY	DAKYSVKVSF	GGGQYDLPPW	SISILPDCKT	429
carnation	385	----------	GSCAAFLANY	DPKWSVKVTF	SGMEFELPAW	SISILPDCKK	434
asparagus	381	----------	-SCAAFLANF	NSRYYATVTF	NGMHYNLPPW	SVSILPDCKT	430
broccoli	381	S---------	-SC--FIGNV	NATADALVNF	KGKDYNVPAW	SVSVLPDCDK	430
Lupin	389	----------	A-CAAFLANY	NTDYSTQVKF	GNGQYDLPPW	SISILPDCKT	438

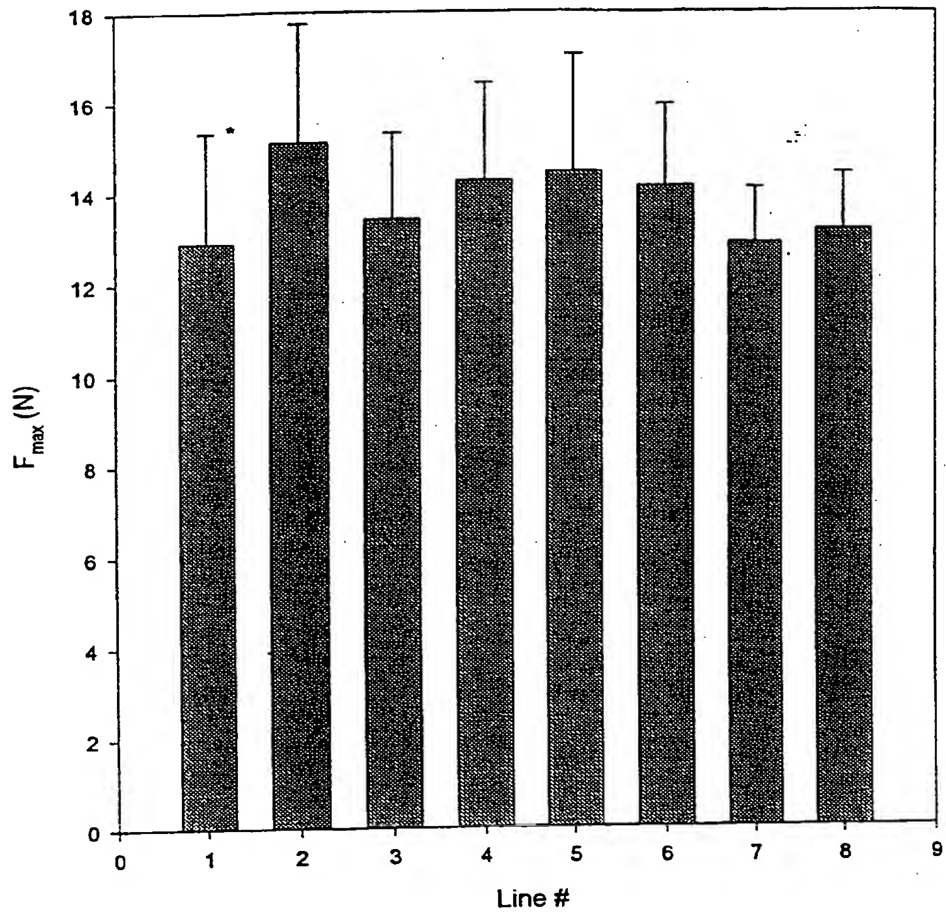
		460	470	480	490	500	
TBG1-ORF	427	TVYNTARVGA	QSAQM--K--	----------	-------MTP	VSRGFS--WE	476
TBG2-ORF	437	AEIQLSTQLR	WGHKLQSKQW	AQILFQLGII	LCFYKLSLKA	SSESFSQSWM	486
TBG3-ORF	431	TVFNTARIGA	QSAQM--K--	----------	-------MTP	VSRGLP--WQ	480
TBG4-ORF	429	AVYNTAQVNS	QSSSI--K--	----------	-------MTP	AGGGLS--WQ	478
TBG5-ORF	451	----------	----------	----------	----------	----------	500

```
DNASIS
Multiple  Edit1
                                   Figure 3
                                   Sheet 3 of 4
                                                                                   500
TBG6-ORF     451 ---------- ---------- ---------- ---------- ----------           499
TBG7-ORF     450 VAFNTAKVGC QTSIVNMAP- --------ID L--HPTASSE KRDIKSLQWE           479
apple        430 EVYNTAKVGS QSSQV--Q-- ---------- -------MTP VHSGEP--WQ           484
carnation    435 EVYNTARVNE PSPKLHSK-- ---------- -------MTP VISNLN--WQ           480
asparagus    431 TVFNTARVGA QTTTM--K-- ---------- -------MQY LG-GES--WK           480
broccoli     431 EAYNTARVNT QTSIITEDS- ---------- -G-------D EPEKLKWTWR           480
Lupin        439 EVFNTAKVNS PRLHR--K-- ---------- -------MTP VNSAFA--WQ           488

                        510        520        530        540        550
TBG1-ORF     477 S-FNEDAASH EDD-TSIVVG LLEQINITRD VSDYLWYMTD IEIDPTE-GF           526
TBG2-ORF     487 T-LKEPLGVW GDKN-FTSKG ILEHLNVTKD QSDYLWYLTR IYISDDDISF           536
TBG3-ORF     481 S-FNEETSSY EDS-SFTVVG LLEQINTTRD VSDYLWYSTD VKIDSRE-KF           530
TBG4-ORF     479 S-YNEETPTA DDSDTLTANG LWEQKNVTRD SSDYLWYMTN VNIASNE-GF           528
TBG5-ORF     501 ---------- ---------- ---------- ---------- ----------           550
TBG6-ORF     501 ---------- ---------- ---------- ---------- ----------           550
TBG7-ORF     500 V-FKETAGVW GVAD-ETKNG FVDHINTTKD ETDYLWYTTS IFVHAEE-DE           549
apple        480 S-FIEETTSS DETDTTTLDG LYEQINITRD TTDYLWYMTD ITIGSDE-AE           529
carnation    485 S-YSDEVPTA DSPGTSREKK LYEQINMTWD KSDYLWYMTD VVLDGNE-GF           534
asparagus    481 A-YTEDTDAL NDN-TSTKDG LVEQLSTTWD RSDYLWYTTY VDIAKNE-EE           530
broccoli     481 PEFTTQKTIL KGSGDLIAKG LVDQKDVTND ASDYLWYMTR VHLDKKDPIW           530
Lupin        489 S-YNEEPASS SENDPVTGYA LWEQVGVTRD SSDYLWYLTD VNIGPND---           538

                        560        570        580        590        600
TBG1-ORF     527 LNSGN-WPWL TVFSAGHALH VFVNGQLAGT VYGSLENPKL TFSNGINLRA           576
TBG2-ORF     537 WEENDVSPTI DIDSMRDFVR IFVNGQLAGS VKGKW----I KVVQPVKLVQ           586
TBG3-ORF     531 LRGGK-WPWL TIMSAGHALH VFVNGQLSGT AYGSLEKPKL TFSKAVNLRA           580
TBG4-ORF     529 LKNGK-DPYL TVMSAGHVLH VFVNGKLSGT VYGTLDNPKL TYSGNVKLRA           578
TBG5-ORF     551 ---------- ---------- ---------- ---------- ----------           600
TBG6-ORF     551 ---------- ---------- ---------- ---------- ----------           600
TBG7-ORF     550 LRN-RGTAML FVESKGHAMH VFINKKLQAS ASGNGTVPQF KEGTPIALKA           599
apple        530 LKNGK-SPLL TIFSAGHALN VFINGQLSGT VYGSLENPKL SFSQNVNLRS           579
carnation    535 LKKGD-EPWL TVNSAGHVLH VFVNGQLQGH AYGSLAKPQL TFSQKVKMTA           584
asparagus    531 LKTGK-YPYL TVMSAGHAVH VFINGQLSGT AYGSLDNPKL TYSGSAKIWA           580
broccoli     531 SRNMS----L RVHSNAHVLH AYVNGKYVGN QIVRDNKFDY RFEKKVNLVH           580
Lupin        539 IKDGK-WPVL TAMSAGHVLN VFINGQYAGT AYGSLDDPRL TFSQSVNLRV           588

                        610        620        630        640        650
TBG1-ORF     577 GVNKISLLSI AVGLPNVGPH FETWNAGVLG PVSLNGLNEG T---RDLTWQ           626
TBG2-ORF     587 GYNDILLLSE TVGLQNYGAF LEKDGAGFKG QIKLTGCKSG D---INLTTS           636
TBG3-ORF     581 GVNKISLLSI AVGLPNIGPH FETWNAGVLG PVSLTGLDEG K---RDLTWQ           630
TBG4-ORF     579 GINKISLLSV SVGLPNVGVH YDTWNAGVLG PVTLSGLNEG S---RNLAKQ           628
TBG5-ORF     601 ---------- ---------- ---------- ---------- ----------           650
TBG6-ORF     601 ---------- ---------- ---------- ---------- ----------           650
TBG7-ORF     600 GKNEISLLSM TVGLQTAGAF YE-WIGAGPT SVKVAGFKTG T---MDLTAS           649
apple        580 GINKLALLSI SVGLPNVGTH FETWNAGVLG PITLKGLNSG T---WDMSGW           629
carnation    585 GVNRISLLSA VVGLANVGWH FERYNQGVLG PVTLSGLNEG T---RDLTWQ           634
asparagus    581 GSNKISILSV SVGLPNVGNH FETWNTGVLG PVTLTGLNEG K---RDLSLQ           630
broccoli     581 GTNHLALLSV SVGLQNYGPF FESGPTGING PVKLVGYKGD ETIEKDLSKH           630
Lupin        589 GNNKISLLSV SVGLANVGTH FETWNTGVLG PVTLTGLSSG T---WDLSKQ           638

                        660        670        680        690        700
TBG1-ORF     627 KWFYKVGLKG EALSLHSLSG SPSVE--WVE GSLVAQKQPL SWYKTTFNAP           676
TBG2-ORF     637 LWTYQVGLRG EFLEVYDVNS TESAG--WTE FPTGTTPSVF SWYKTKFDAP           686
TBG3-ORF     631 KWSYKVGLKG EALSLHSLSG SSSVE--WVE GSLVAQRQPL TWYKSTFNAP           680
TBG4-ORF     629 KWSYKVGLKG ESLSLHSLSG SSSVE--WVR GSLMAQKQPL TWYKATFNAP           678
TBG5-ORF     651 ---------- ---------- ---------- ---------- ----------           700
TBG6-ORF     651 ---------- ---------- ---------- ---------- ----------           700
TBG7-ORF     650 AWTYKIGLQG EHLRIQKSYN LKSKI--WAP TSQPPKQQPL TWYKAVVDAP           699
apple        630 KWTYKTGLKG EALGLHTVTG SSSVE--WVE GPSMAEKQPL TWYKATFNAP           679
carnation    635 YWSYKIGTKG EEQQVYNSGG SSHVQ--WGP PAW---KQPL VWYKTTFDAP           684
asparagus    631 KWTYQIGLHG ETLSLHSLTG SSNVE--WGE AS---QKQPL TWYKTFFNAP           680
broccoli     631 QWDYKIGLNG FNHKLFSMKS AGHHHRKWST EKLPADRM-L SWYKANFKAP           680
Lupin        639 KWSYKIGLKG ESLSLHTEAG SNSVE--WVQ GSLVAKKQPL AWYKTTFSAP           688

                        710        720        730        740        750
TBG1-ORF     677 DGNEPLALDM NTMGKGQVWI NGQSLGRHWP AYKSS-GSCS V-CNYTGWFD           726
```

```
DNASIS                                      Figure 3
Multiple  Edit1                             Sheet 4 f 4

TBG2-ORF     687  GGTDPVALDF SSMGKGQAWV NGHHVGRYWT LVAPN-NGGG RTCDYRGAYH   736
TBG3-ORF     681  AGNDPLALDL NTMGKGQVWI NGQSLGRYWP GYKAS-GNCG A-CNYAGWFN   730
TBG4-ORF     679  GGNDPLALDM ASMGKGQIWI NGEGVGRHWP GYIAQ-GDCS K-CSYAGTFN   728
TBG5-ORF     701  ---------- ---------- ---------- ---------- ----------   750
TBG6-ORF     701  ---------- ---------- ---------- ---------- ----------   750
TBG7-ORF     700  PGNEPVALDM IHMGKGMAWL NGQEIGRYWP RRTSKYENCV TQCDYRGKFN   749
apple        680  PGDAPLALDM GSMGKGQIWI NGQSVGRHWP GYIAR-GSCG D-CSYAGTYD   729
carnation    685  GGNDPLALDL GSMGKGQAWI NGQSIGRHWS NNIAK-GSCN DNCNYAGTYT   734
asparagus    681  PGNEPLALDM NTMGKGQIWI NGQSIGRYWP AYKAS-GSCG S-CDYRGTYN   730
broccoli     681  LGKDPVIVDL NGLGKGEVWI NGQSIGRYWP SFNSSDEGCT EECDYRGEYG   730
Lupin        689  AGNDPLALDL GSMGKGEVWV NGQSIGRHWP GNKAR-GNCG N-CNYAGTYT   738

                         760        770        780        790        800
TBG1-ORF     727  EKKCLTNCGE GSQRWYHVPR SWLYPTGNLL V-VFEEWGGD PYGITLVKRE   776
TBG2-ORF     737  SDKCRTNCGE ITQAWYHIPR SWLKTLNNVL V-IFEETDKT PFDISISTRS   786
TBG3-ORF     731  EKKCLSNCGE ASQRWYHVPR SWLYPTGNLL V-LFEEWGGE PHGISLVKRE   780
TBG4-ORF     729  EKKCQTNCGQ PSQRWYHVPR SWLKPSGNLL V-VFEEWGGN PTGISLVRRS   778
TBG5-ORF     751  ---------- ---------- ---------- ---------- ----------   800
TBG6-ORF     751  ---------- ---------- ---------- ---------- ----------   800
TBG7-ORF     750  PDKCVTGCGQ PTQRWYHVPR SWFKPSGNVL I-IFEEIGGD PSQIRFSMRK   799
apple        730  DKKCRTHCGE PSQRWYHIPR SWLTPTGNLL V-VFEEWGGD PSRISLVERG   779
carnation    735  ETKCLSDCGK SSQKWYHVPR SWLQPRGNLL V-VFEEWGGD TKWVSLVKRT   784
asparagus    731  EKKCLSNCGE ASQRWYHVPR SWLIPTGNFL V-VLEEWGGD PTGISMVKRS   780
broccoli     731  SDKCAFMCGK PTQRWYHVPR SFLNDKGHNT ITLFEEMGGD PSMVKFKTVV   780
Lupin        739  DTKCLANCGQ PSQRWYHVPR SWLRSGGNYL V-VLEEWGGD PNGIALVERT   788

                         810        820        830        840        850
TBG1-ORF     777  IGSVCADIYE WQ-PQLLNWQ RLVSGKFDRP LR--PKAHLK CAPGQKISSI   826
TBG2-ORF     787  TETICAQVSE KHYPPLHKWS HSEFDRKLSL MDKTPEMHLQ CDEGHTISSI   836
TBG3-ORF     781  VASVCADINE WQ-PQLVNWQ MQASGKVDKP LR--PKAHLS CASGQKITSI   830
TBG4-ORF     779  ---------- ---------- --------R- ---------- ----------   828
TBG5-ORF     801  ---------- ---------- ---------- ---------- ----------   850
TBG6-ORF     801  ---------- ---------- ---------- ---------- ----------   850
TBG7-ORF     800  VSGACGHLSV -DHPSFD--V ENLQGSEIEN DKNRPTLSLK CPTNTNISSV   849
apple        780  ---------- ---------- --------TA LD--AK---- ----------   829
carnation    785  IR-------- ---------- ---------- -R--PKVHLS CDPGQKMSKI   834
asparagus    781  VASVCAEVEE LQ-PTMDNWR TKAYG----- ---------- ----------   830
broccoli     781  TGRVCAKAHE ---------- ---------- ---HNKVELS CN-NRPISAV   830
Lupin        789  ---------- ---------- ---------- ---------- ----------   838

                         860        870        880        890        900
TBG1-ORF     827  KFASFGTPEG VCGNFQQGSC HAPRSYDAFK K-----NCVG KESCSVQVTP   876
TBG2-ORF     837  EFASYGSPNG SCQKFSQGKC HAANSLSV-- ---VSQACIG RTSCSIGISN   886
TBG3-ORF     831  KFASFGTPQG VCGSFREGSC HAFHSYDAFE R-----YCIG QNSCSVPVTP   880
TBG4-ORF     829  ---------- ---------- ---------- ---------- ----------   878
TBG5-ORF     851  ---------- ---------- ---------- ---------- ----------   900
TBG6-ORF     851  ---------- ---------- ---------- ---------- ----------   900
TBG7-ORF     850  KFASFGNPNG TCGSYMLGDC HDQNSAALVE K-----VCLN QNECALEMSS   899
apple        830  ---------- ---------- ---------- ---------- ----------   879
carnation    835  ---------- ---------- ---------- ---------- ----------   884
asparagus    831  KFASFGTPQG TCGSFSEGSC HAHKSYDAFE QEGLMQNCVG QEFCSVNVAP   880
broccoli     831  KFASFGNPSG QCGSFAAGSC EGAKDAVKV- ---VAKECVG KLNCTMNVSS   880
Lupin        839  ---------- ---------- ---------- ---------- ----------   888

                         910        920        930        940        950
TBG1-ORF     877  ENFGGDP-CR NVLKKLSVEA ICS------- ----...... ..........   926
TBG2-ORF     887  GVFG-DP-CR HVVKSLAVQA KCSPPPDLST SASS...... ..........   936
TBG3-ORF     881  EIFGGDP-CP HVMKKLSVEV ICS------- ----...... ..........   930
TBG4-ORF     879  ---------- ---------- ---------- ----...... ..........   928
TBG5-ORF     901  ---------- ---------- ---------- ----...... ..........   950
TBG6-ORF     901  ---------- ---------- ---------- ----...... ..........   950
TBG7-ORF     900  ANFNMQL-CP STVKKLAVEV NCS------- ----...... ..........   949
apple        880  ---------- ----KL---- ---------- ----...... ..........   929
carnation    885  ---------- ---------- ---------- ----...... ..........   934
asparagus    881  EVFGGDP-CP GTMKKLAVEA ICE------- ----...... ..........   930
broccoli     881  HKFGSNLDCG DSPKRLFVEV EC-------- ----...... ..........   930
```

Figure 11D

4-mm cylindrical puncture to [redacted]mm
Breaker + 7 8



* Standard Deviation

PU07 Line#	PU07 Mean	PU07 Std Dev
1	12.91	2.43
5	15.13	2.61
6	13.44	1.90
7	14.28	2.16
8	14.47	2.58
9	14.14	1.81
11	12.90	1.20
12	13.18	1.25

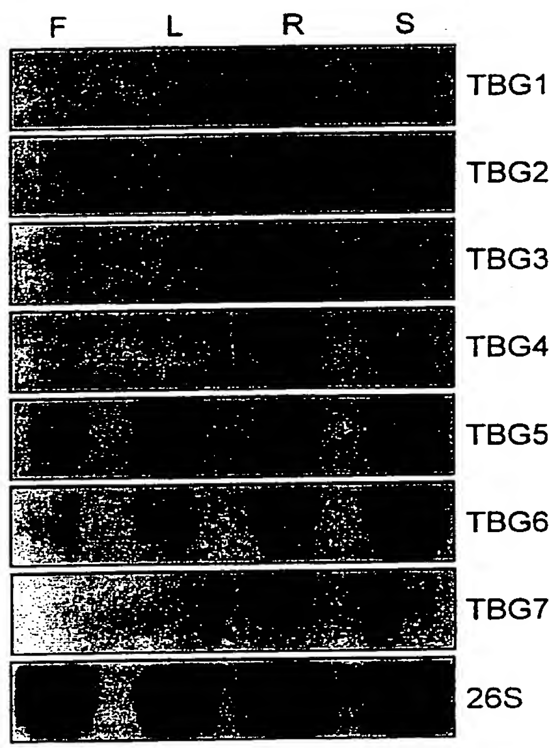


Figure 4. Autoradiograph of northern blot analysis of TBG expression in various plant tissues. Twenty µg of total RNA extracted from flowers (F), leaves (L), roots (R) and stems (S) was loaded in each lane. RNAs were separated in an agarose gel and transferred to nylon membrane. Blots were hybridized using the probes indicated to the right, washed to a final stringency of 0.1X SSC at 65°C and were used to expose x-ray film. A 26S ribosomal gene clone from soybean was used as a loading control for each blot and one example is shown.

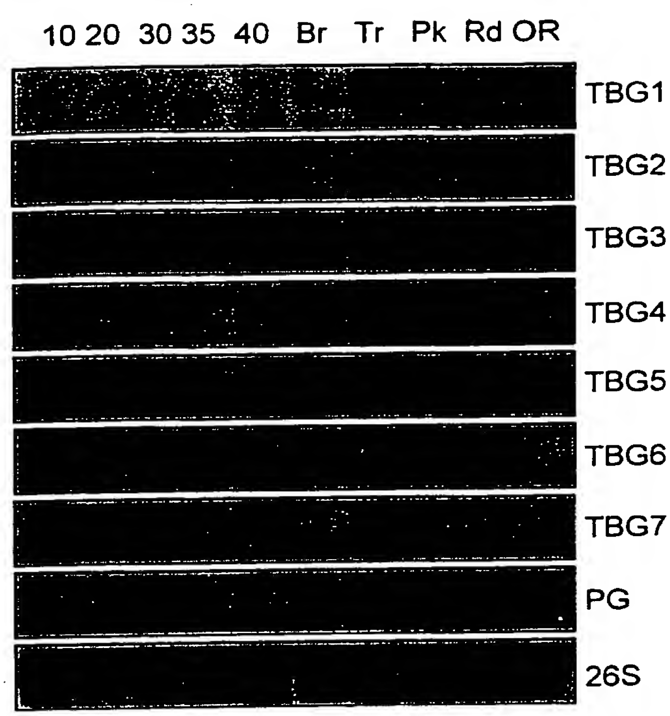


Figure 5. Autoradiograph of northern blot analysis of TBG expression in fruit tissues. Twenty μg of total RNA extracted from peel and outer pericarp tissue was loaded in each lane. Fruit were harvested at 10, 20, 30, 35, and 40 days post-pollenation and at the breaker (Br), turning (Tr), pink (Pk), red (Rd) and over ripe (OR) stages. RNAs were separated in an agarose gel and transferred to nylon membrane. Blots were hybridized using the probes indicated to the right, washed to a final stringency of 0.1X SSC at 65°C and were used to expose x-ray film. A 26S ribosomal gene clone from soybean was used as a loading control for each blot and one example is shown. A cDNA clone for tomato polygalacturonase (PG) was also used as a probe to show a well characterized, fruit-ripening-specific control.

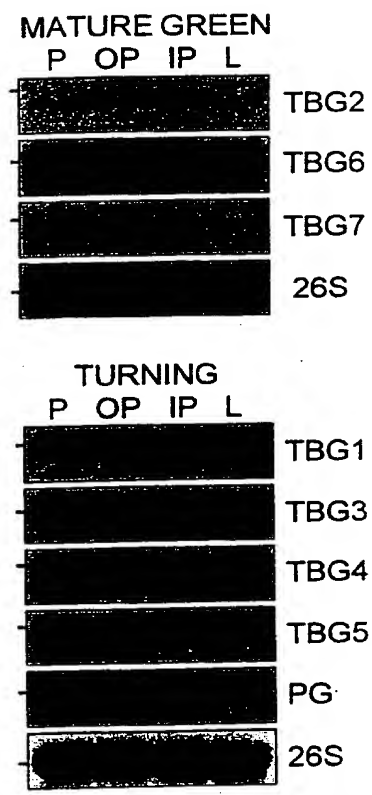


Figure 6. Autoradiograph of northern blot analysis of TBG expression in fruit tissues. Twenty μg of total RNA extracted from mature green or turning stage fruit peel (P), outer pericarp (OP), inner pericarp (IP) and locular (L) tissue was loaded in each lane. RNAs were separated in an agarose gel and transferred to nylon membrane. Blots were hybridized using the probes indicated to the right, washed to a final stringency of 0.1X SSC at 65°C and were used to expose x-ray film. A 26S ribosomal gene clone from soybean was used as a loading control for each blot and one example is shown. A cDNA clone for tomato polygalacturonase (PG) was also used as a probe to show a well characterized, fruit-ripening-specific control.

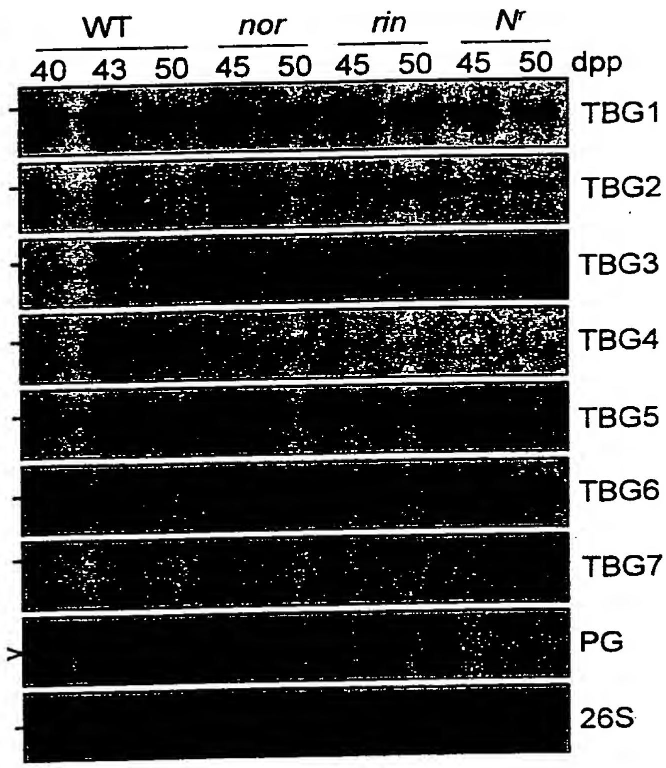


Figure 7. Autoradiograph of northern blot analysis of TBG expression in normal and mutant fruit tissues. Twenty μg of total RNA extracted from peel and outer pericarp tissue at various days post-pollination (dpp) was loaded in each lane. RNAs were separated in an agarose gel and transferred to nylon membrane. Blots were hybridized using the probes indicated to the right, washed to a final stringency of 0.1X SSC at 65°C and were used to expose x-ray film. A 26S ribosomal gene clone from soybean was used as a loading control for each blot and one example is shown. A cDNA clone for tomato polygalacturonase (PG) was also used as a probe to show a well characterized, fruit-ripening-specific control. The - and > marks on the left indicate the position of the tomato 27S and 18S rRNAs respectively.

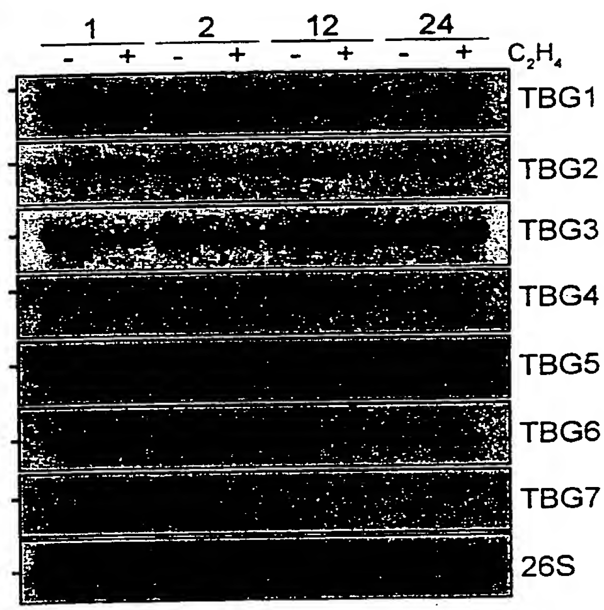


Figure 8. Autoradiograph of northern blot analysis of TBG expression in response to ethylene treatment of mature green fruit tissues. Twenty µg of total RNA extracted from peel and outer pericarp tissue at various times (1, 2, 12 and 24 hours) after treatment with (+) or without (-) 10 ppm ethylene was loaded in each lane. RNAs were separated in an agarose gel and transferred to nylon membrane. Blots were hybridized using the probes indicated to the right, washed to a final stringency of 0.1X SSC at 65°C and were used to expose x-ray film. A 26S ribosomal gene clone from soybean was used as a loading control for each blot and one example is shown. The - marks on the left indicate the position of the tomato 27S rRNA.

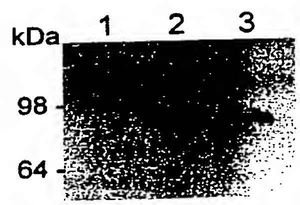


Figure 9. Western blot analysis of TBG4 expression by yeast. A yeast clone was isolated that secreted high levels of FLAG-TBG4 fusion protein into the culture medium. Protein samples were separated in an 8% acrylamide gel, transferred to nitrocellulose and were blotted with M1 anti-FLAG primary antibody. Blots were washed and blotted with an alkaline-phosphatase conjugated secondary antibody and alkaline phosphatase activity was detected using Sigma Fast substrate. Lane 1, culture medium of an untransformed yeast clone was used as a negative control. Lane 2, culture medium of yeast clone expressing FLAG-TBG4 fusion protein. Lane 3, Affinity purified FLAG-TBG4 fusion protein.

Figure 10

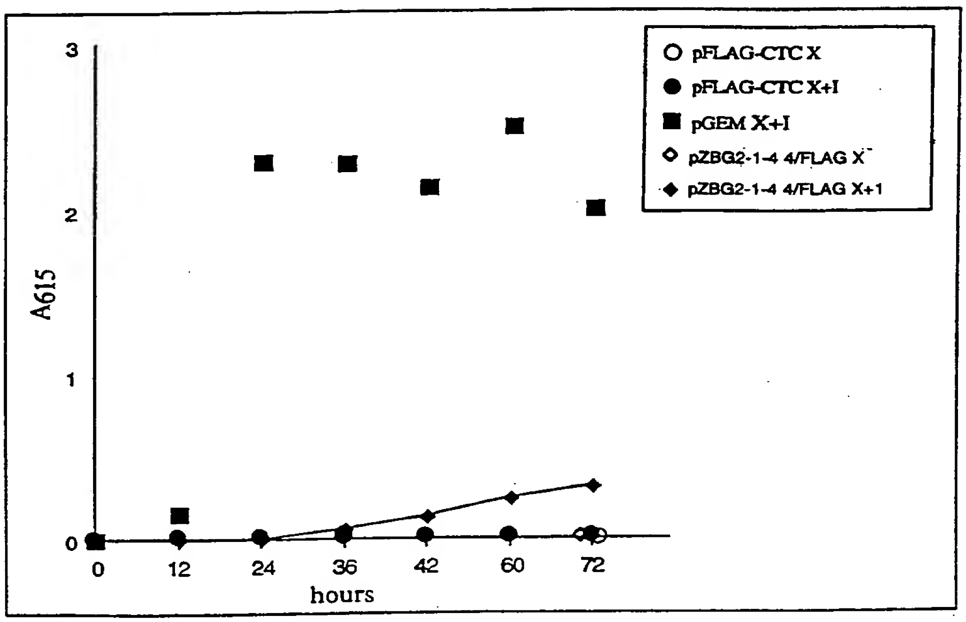
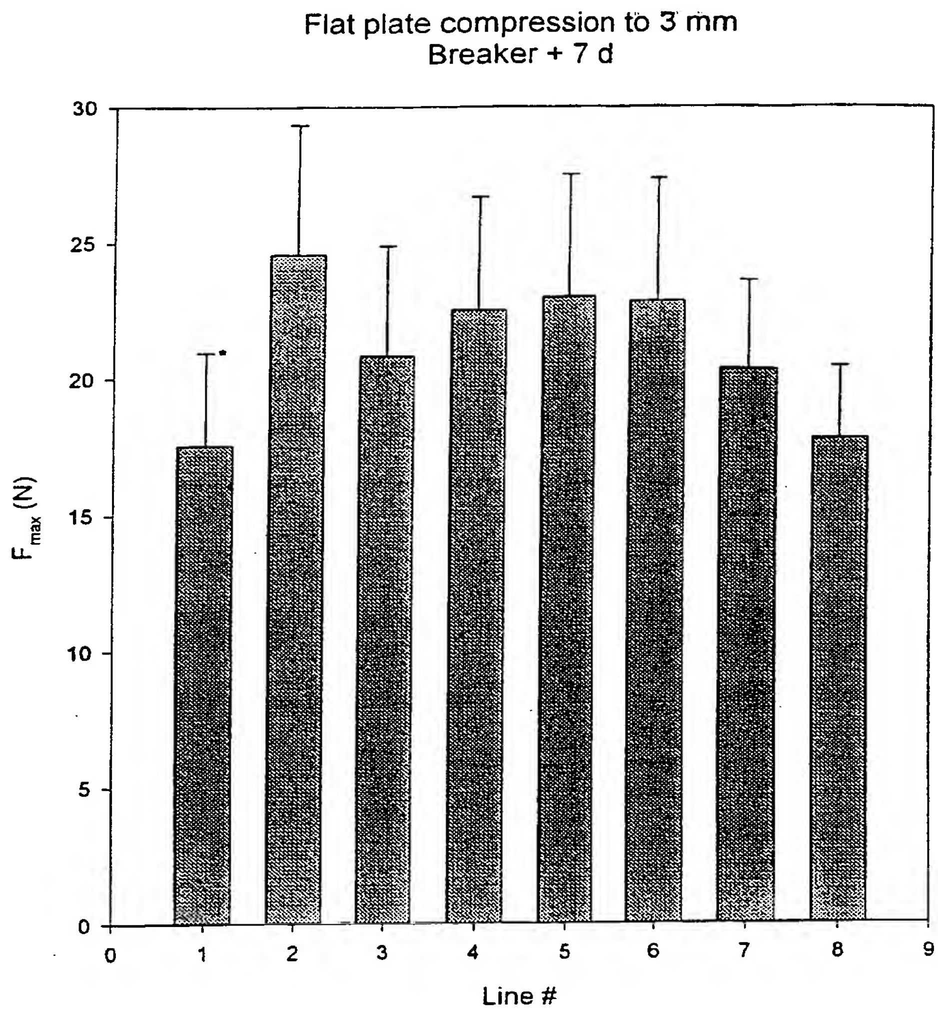

pFLAG-CTC X
pFLAG-CTC X+I
pGEM X+I
pZBG2-1-4 4/FLAG X
pZBG2-1-4 4/FLAG X+1
A615
3
2
1
0
0
12
24
36
42
60
72
hours

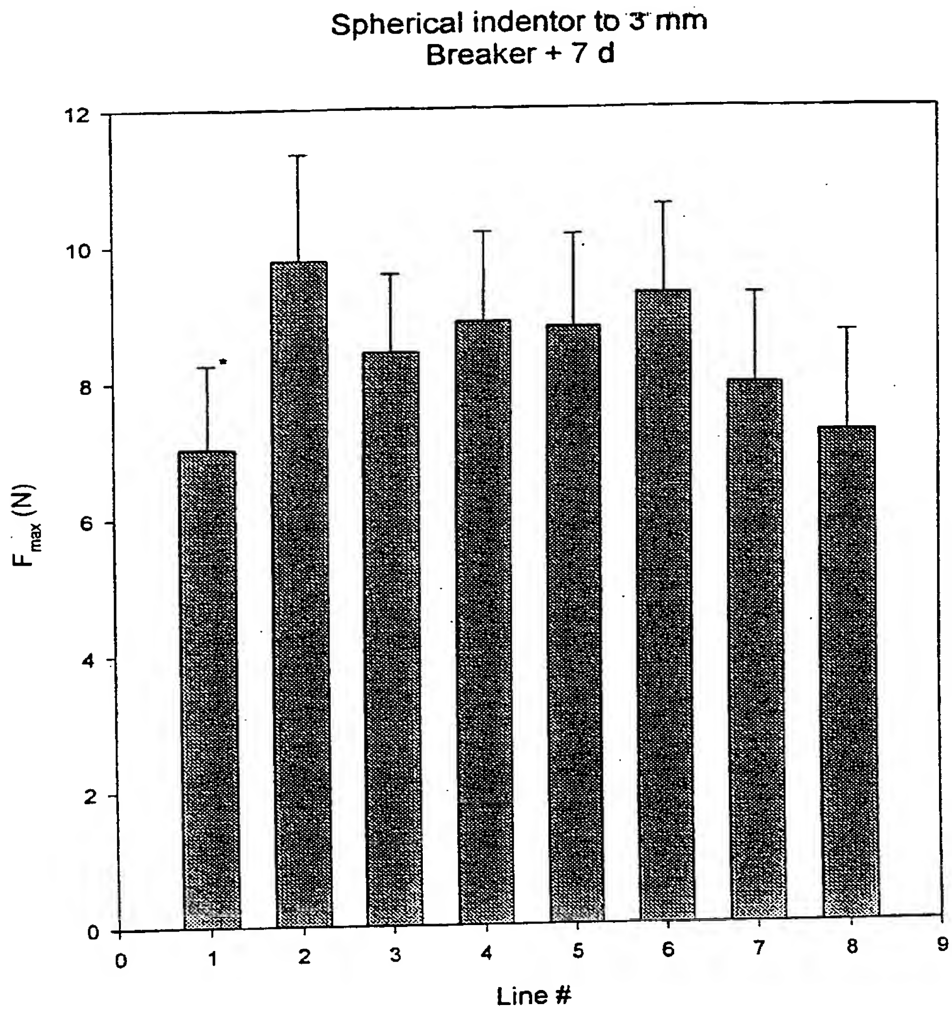
Figure 11A



* Standard Deviation

FP07 Line #	FP07 mean	FP07 std dev
1	17.52665	3.418542
2	24.56026	4.786548
3	20.81681	4.066194
4	22.54655	4.15923
5	23.03255	4.493091
6	22.84338	4.517462
7	20.36124	3.24608
8	17.81924	2.665468

Figure 11B

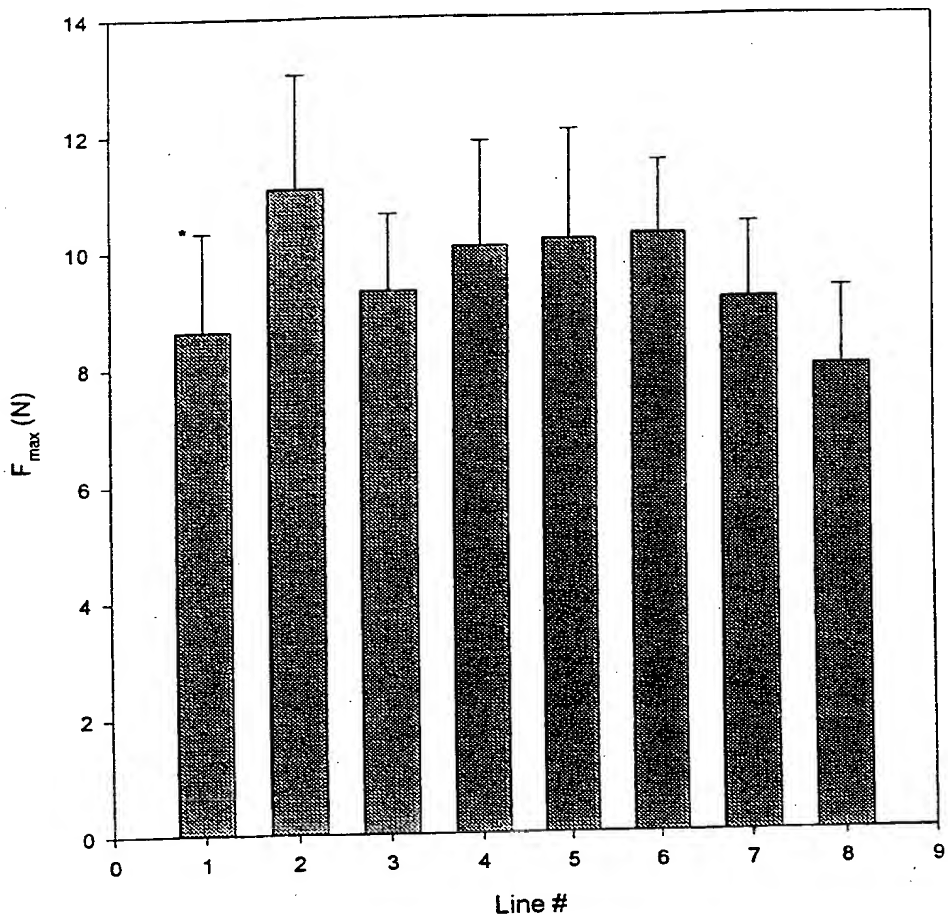


* Standard Deviation

SP07 Line #	SP07 Mean	SP07 Std Dev
1	7.02	1.22
5	9.77	1.57
6	8.43	1.15
7	8.87	1.32
8	8.78	1.36
9	9.28	1.29
11	7.96	1.30
12	7.26	1.45

Figure 11C

4-mm cylindrical indentor to [redacted] mm
Breaker + 7 d



* Standard Deviation

CY07 LINE#	CY07 Mean	CY07 Std Dev
1	8.62	1.69
5	11.07	1.96
6	9.31	1.33
7	10.07	1.81
8	10.18	1.88
9	10.27	1.26
11	9.15	1.30
12	7.99	1.33

FI[illegible]1 E(1)
Flat plate compression to 3 mm
Breaker + 7 d

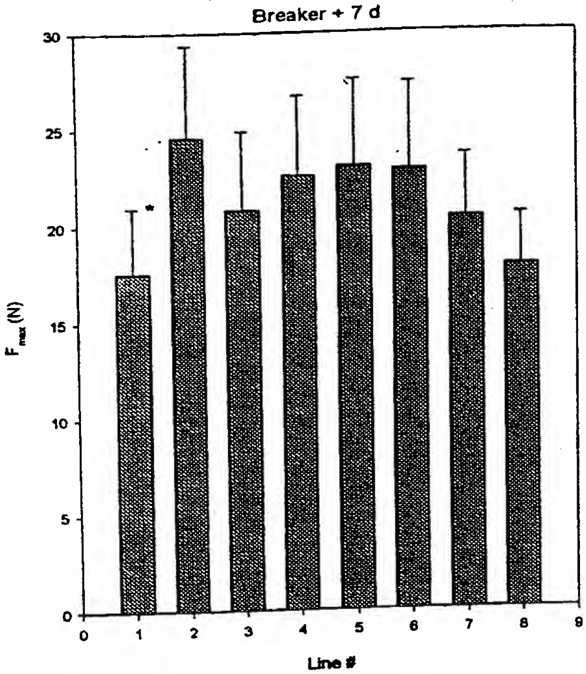


FIG 11 E (2)
Spherical indentor to 3 mm
Breaker + 7 d

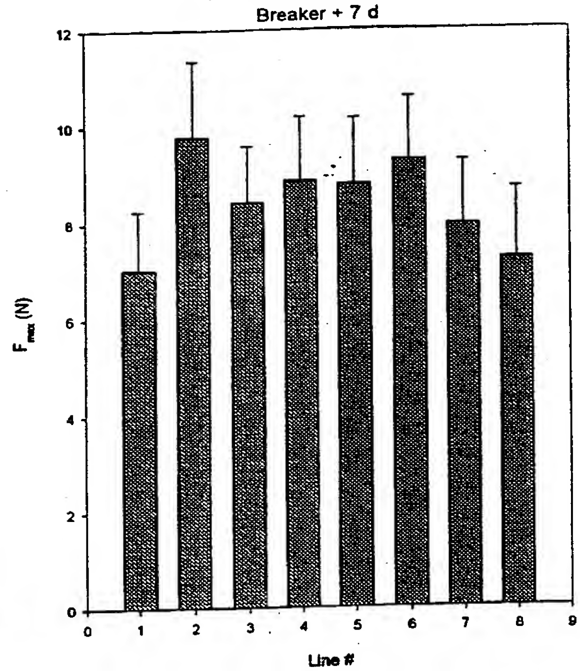


FIG 11 E(3)
4-mm cylindrical indentor to 3 mm
Breaker + 7 d

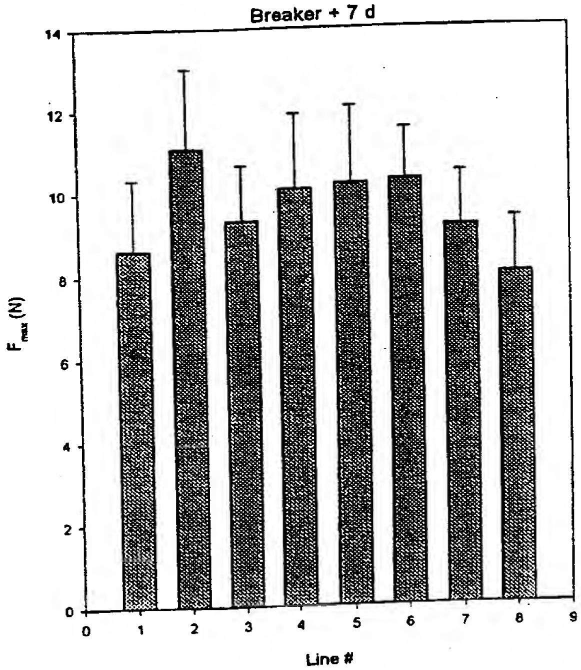
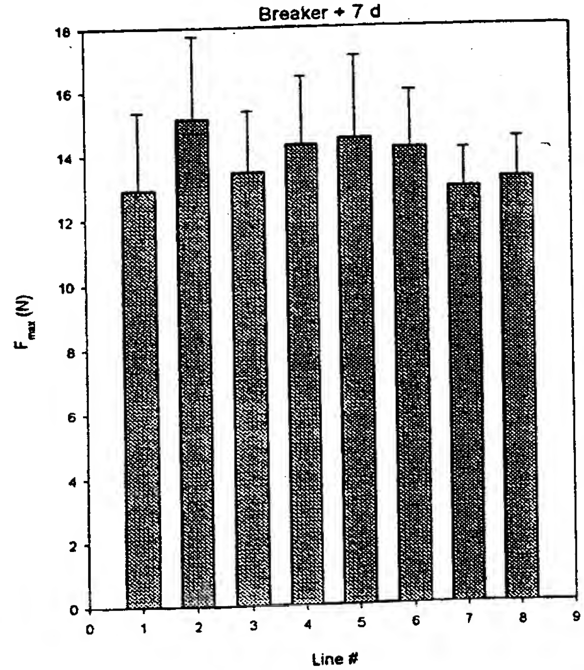
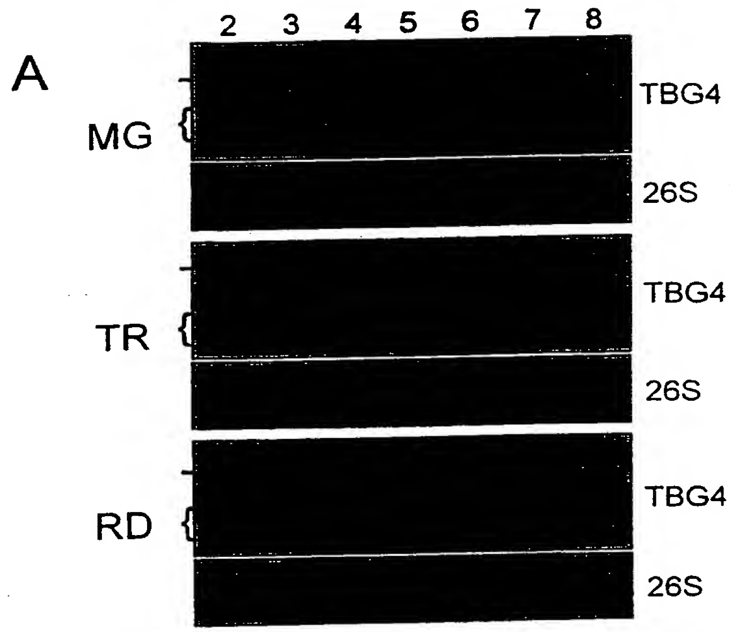


FIG. 11 E (4)
4-mm cylindrical puncture to 10 mm
Breaker + 7 d



* Standard Deviation



B

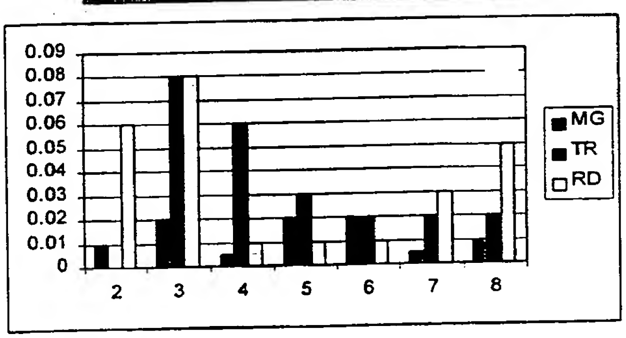


Figure 12. Northern blot analysis of TBG4 expression in transgenic fruit containing TBG4 antisense construct. A. Total RNA was extracted from mature green/42 days post-pollenation (MG), turning/breaker + 3 (TR) and red/breaker + 7 (RD) fruit and twenty μg was loaded in each lane. RNAs were separated in an agarose gel and transferred to nylon membrane. Blots were hybridized using the probes indicated to the right, washed to a final stringency of 0.1X SSC at 65°C and were used to expose x-ray film. A 26S ribosomal gene clone from soybean was used as a loading control. The marks - and { denote the positions of the endogenous TBG4 and antisense mRNAs respectively. Lanes 2-8 correspond to transgenic lines 2-8 in Figures 11A-E. **B.** Chart of TBG4 mRNA levels in lines 2-8. Autoradiographs were scanned using a densitometer and TBG4 mRNA levels were corrected against the loading controls. TBG4 mRNA levels are shown in arbitrary units.

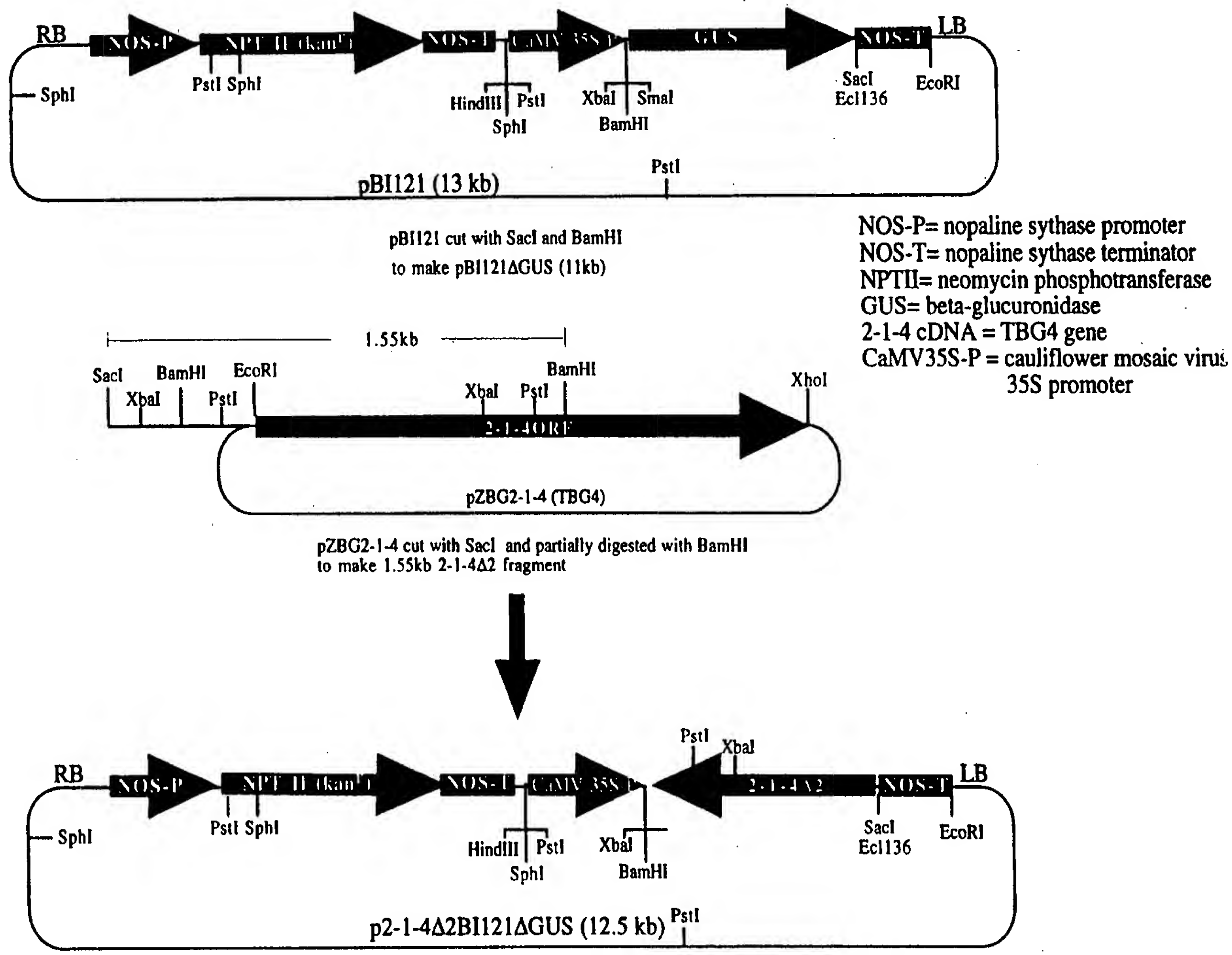


Figure 13. Binary construct used to transform plants and express TBG4 (pZBG2-1-4) in the antisense orientation.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12697

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12N 5/04, 9/38, 15/09, 15/56; A01H 5/00, 5/10
US CL : 435/207, 419, 468; 800/278, 295, 298
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/207, 419, 468; 800/278, 295, 298

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST, CAPLUS, AGRICOLA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SMITH et al. A Gene Coding for Tomato Fruit β-Galactosidase II Is Expressed during Fruit Ripening. Plant Physiology. 1998, Vol. 117, pages 417-423, especially 422-423.	27
Y	ALI et al. Isolation, Characterization and Significance of Papaya β-Galactanases to Cell Wall Modification and Fruit Softening during Ripening. Physiologia Plantarum. 1998, Vol. 104, pages 105-115, especially page 111, col. 2, and page 113, col. 2.	27

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 OCTOBER 1999	03 NOV 1999

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	MELISSA KIMBALL JOYCE BRIDGERS PARALEGAL SPECIALIST CHEMICAL MATRIX
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12697

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-26 and 28-32
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

because the claims all recite SEQ ID No.s or depend therefrom while no CRF has been filed for this case. Therefore it is not possible to search the claimed nucleic acid and amino acids nor is it possible to search transgenic seeds or plants comprising the sequences.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12697

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CARRINGTON et al. β-Galactosidase II Activity in Relation to Changes in Cell Wall Galactosyl Composition during Tomato Ripening. Journal of the American Society of Horticultural Science. 1996, Vol. 121, No. 1, pages 132-136, especially page 135, col. 2.	27
Y	PRESSEY, R. β-Galactosidases in Ripening Tomatoes. Plant Physiology. 1983, Vol. 71, pages 132-135, see entire article.	27
Y,P	US 5,859,344 A (BIRD et al.) 12 January 1999, see entire document.	27